Vol. XXXIII. Part II August, 1944

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

KARL PEARSON

ASSISTED BY

EGON S. PEARSON

Reprinted by offset-litho

ISSUED BY THE BIOMETRIC LABORATORY
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

VOLUME XXXIII, PART II AUGUST 1944

ON AUTOREGRESSIVE TIME SERIES

BY M. G. KENDALL

1. The product-moment correlation coefficients obtained by correlating the members of a time-series among themselves provide a useful method of investigating the behaviour of the series, especially in regard to oscillatory movements. Consider a series of values $x_1 \dots x_n$, measured about their mean, trend-free and defined at equal intervals of time $t = 1, \dots, n$. The kth serial correlation r_k is defined as the correlation between members of the series k intervals apart,

$$r_k = \frac{\sum_{j=1}^{n-k} x_j x_{j+k}}{\left\{\sum_{j=1}^{n-k} x_j^2 \sum_{j=1}^{n-k} x_{j+k}^2\right\}^{\frac{1}{2}}}. \tag{1}$$

The figure obtained by plotting r_k as ordinate against k as abscissa and joining each point to the next is known as the correlation diagram or correlogram. Since $r_0 = 1$ the correlogram always starts from the point (0, 1). When the series is infinite in extent, the *serial* correlations become *auto*-correlations (denoted by a Greek ρ), and the series for which such correlations do not vanish may be said to be autocorrelated.

2. If the time series is random the correlogram presents no systematic appearance. If it consists of a simple harmonic the correlogram reproduces that harmonic. If it results from a moving average of finite extent d of a random series the correlogram will, within sampling limits, be zero after $k = d$. The correlogram thus provides a criterion for distinguishing between various kinds of oscillatory time series. In this paper I propose to consider the correlogram of a series defined by the difference equation

$$u_{t+2} + a u_{t+1} + b u_t = \epsilon_{t+2}, \tag{2}$$

where a and b are constants and ϵ_{t+2} is a random variable. This equation I shall call the generating equation and the coefficients a and b generating coefficients. The generated series may be said to be autoregressive. The notion of generating series in this way was introduced by Yule (1927) in a classical paper on sunspot periodicities and has been applied to meteorological and economic series with some success. The autoregressive scheme can, in fact, explain a typical phenomenon of such series for which Fourier and periodogram analysis cannot satisfactorily account without great artificiality, namely, the continual shift in phase and variation of amplitude which occur even when the series is smooth. Accounts of the autoregressive scheme have been given in the books by Wold (1938) and Davis (1941).

3. The equation (2) may be solved by the ordinary methods appropriate to difference equations. If the roots of

$$\xi^2 + a\xi + b = 0$$

are $\alpha + i\beta$, $\alpha - i\beta$, the complementary function of (2) is

$$p^t(A \cos \theta t + B \sin \theta t), \tag{3}$$

where $p = +\sqrt{b}$, $\theta = \arctan\frac{\beta}{\alpha} = \arctan\sqrt{\left(\frac{4b}{a^2} - 1\right)}$, and A and B are arbitrary constants.

It is here assumed that b is positive and that $4b > a^2$. It is also to be assumed that $\sqrt{b} = p$ is not greater than unity. The complementary function (3) then represents a damped harmonic. I shall call $2\pi/\theta$ the fundamental period of the generated system.

Let ξ_t be a particular value of (3) such that

$$\xi_0 = 0, \quad \xi_1 = 1, \tag{4}$$

i.e. such that

$$\xi_t = \frac{2}{\sqrt{(4p^2 - a^2)}}\, p^t \sin \theta t. \tag{5}$$

Then a particular integral of (2) will be found to be

$$\sum_{j=0}^{\infty} \xi_j \epsilon_{t-j+1}. \tag{6}$$

The complete solution is

$$u_t = p^t (A \cos \theta t + B \sin \theta t) + \sum_{j=0}^{\infty} \xi_j \epsilon_{t-j+1}. \tag{7}$$

In practical cases we may assume that the series was 'started up' some time ago, so that the complementary function has been damped out of existence. The series is then given by

$$u_t = \sum_{j=0}^{\infty} \xi_j \epsilon_{t-j+1}. \tag{8}$$

This is a moving sum of a random series with damped harmonic weights. It has been generally assumed, apparently on the basis of experimental evidence, that the mean period of the generated series will be $2\pi/\theta$, the same as the fundamental period present in the term ξ. This, however, is not necessarily so. Something depends on what we call a period in the generated series, whether, for example, we decide to exclude small ripples on the main wave. One possibility is to define the period as the distance between successive 'upcrosses', i.e. points where the series changes sign from negative to positive. The mean distance between major peaks or major troughs will not, probably, be very different from this in the majority of practical cases.*

Some idea of the mean length of the period as so defined can be obtained for particular distributions of ϵ, though a general discussion presents great difficulties. Consider the sum, which will be required again later,

$$\sum_{j=0}^{\infty} \xi_j \xi_{j+k} = \frac{4}{4p^2 - a^2} \Sigma \{p^{2j+k} \sin \theta j \sin \theta (j+k)\} = \frac{2p^k}{4p^2 - a^2} \Sigma [p^{2j} \{\cos \theta k - \cos \theta (2j+k)\}]$$

$$= \frac{2p^k}{4p^2 - a^2} \left\{ \frac{\cos \theta k}{1 - p^2} - \frac{\cos \theta k - p^2 \cos \theta (k-2)}{1 - 2p^2 \cos 2\theta + p^4} \right\}. \tag{9}$$

We have

$$\Sigma \xi_j \xi_{j+1} / \Sigma \xi_j^2 = \frac{2p \cos \theta}{1 + p^2} = \frac{-a}{1+b} = \cos \phi, \text{ say}. \tag{10}$$

Now if ϵ is normally distributed, the mean period (from one upcross to the next) is $2\pi/\phi$ (Dodd, 1939). Also it is easily seen that $\cos \theta = \dfrac{-a}{2p} = \dfrac{-a}{2\sqrt{b}}$. Thus the period given by the

* Dodd (1939) points out that in series generated by moving averages of random series there sometimes occur oscillations above or below the x-axis which would not be taken into account by counting upcrosses. The difficulty is to decide which of these is not to be ignored as a 'ripple'. I think that for harmonic weights such as are given by the autoregressive scheme the method of upcrosses is satisfactory, as indicated in paragraph 4.

generating equation is not in general the same as the period in the generated series, defined as the mean difference between upcrosses. The ratio of the first to the second is

$$\frac{\arccos\dfrac{a}{1+b}}{\arccos\dfrac{a}{2\sqrt{b}}}. \tag{11}$$

4. One would expect differences of a similar kind when the variation of ϵ is not normal. Experiments indicate, for example, that rectangular variation of ϵ gives much the same periods as normal variation. It does not seem to have been previously remarked that the mean period of the generated series is not that of the fundamental, but part of the explanation is no doubt due to the fact that for ranges of b encountered in practice (say $0{\cdot}5 \leqslant b \leqslant 1{\cdot}0$) the value of the ratio (11) is not very different from unity. In the extreme case $b = 1$ (when the series becomes undamped) the ratio is exactly unity. If $b = 0{\cdot}5$, $a = -1$, the ratio is $1{\cdot}07$. Notwithstanding the theoretical difference exhibited by equation (11), I think we may take it that for most practical purposes a good estimate of the observed period (upcross to upcross) is that given by the generated equation, *if it is known*. Below I give examples of two artificial series of the type of equation (2) which support this conclusion.

5. Equation (9) provides one further interesting item of information, namely, the relationship between the variance of ϵ and the variance of the generated series. We have

$$\operatorname{var} u = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\left(\sum_{j=0}^{\infty}\xi_j\epsilon_{t-j+1}\right)^2 = \sum_{j=0}^{\infty}\xi_j^2 \operatorname{var}\epsilon,$$

cross-product terms in ϵ vanishing since it is a random variable:

$$\frac{\operatorname{var} u}{\operatorname{var}\epsilon} = \Sigma\xi_j^2 = \frac{2}{4p^2-a^2}\left\{\frac{1}{1-p^2} - \frac{1-p^2\cos 2\theta}{1-2p^2\cos 2\theta+p^4}\right\} = \frac{1+b}{(1-b)\{(1+b)^2-a^2\}}.$$

Thus
$$\frac{\operatorname{var}\epsilon}{\operatorname{var} u} = \frac{1-b}{1+b}\{(1+b)^2-a^2\}. \tag{12}$$

In an actual series of 65 terms (referred to in detail below) with $a = -1{\cdot}1$, $b = 0{\cdot}5$, the observed ratio of variances was $0{\cdot}43$. The value given by (12) is $0{\cdot}35$. In another series of the same length for which $a = -1{\cdot}5$, $b = 0{\cdot}9$, the observed ratio was $0{\cdot}04$, the value given by (12) was $0{\cdot}07$. It is difficult to judge how good the agreement is, but to me it appears satisfactory for such short series. It is noticeable that the generated series may have a very much larger variance than the random series on which it is based.

6. Consider now the autocorrelations of the generated series. We have

$$\operatorname{cov}(u_t u_{t+k}) = \lim\frac{1}{n}\sum_t\Big\{\sum_j(\xi_j\epsilon_{t-j+1})\sum_j(\xi_j\epsilon_{t+k-j+1})\Big\} = \operatorname{var}\epsilon\sum_{j=0}^{\infty}\xi_j\xi_{j+k}$$

$$= \operatorname{var}\epsilon\,\frac{2p^k}{4p^2-a^2}\left\{\frac{\cos\theta k}{1-p^2} - \frac{\cos\theta k - p^2\cos\theta(k-2)}{1-2p^2\cos 2\theta+p^4}\right\}.$$

Hence
$$\rho_k = \frac{p^k\{\cos\theta k+\cos\theta(k-2)-2\cos 2\theta\,\cos k\theta+p^2\cos\theta k-p^2\cos\theta(k-2)\}}{(1+p^2)(1-\cos 2\theta)}$$

$$= \frac{p^k\{\sin(k+1)\theta - p^2\sin(k-1)\theta\}}{(1+p^2)\sin\theta}.$$

Writing
$$\tan\psi = \frac{1+p^2}{1-p^2}\tan\theta,$$

we have
$$\rho_k = \frac{p^k(1-2p^2\cos 2\theta+p^4)^{\frac{1}{2}}}{(1+p^2)\sin\theta}\sin(k\theta+\psi) = p^k\frac{\sin(k\theta+\psi)}{\sin\psi}. \tag{13}$$

Apart from the constant factor, ρ_k is thus the product of the damping factor p^k and a harmonic term which has the fundamental period of the generating equation. It is noteworthy that, although the point $\rho_0 = 1$ is a peak at the beginning of the correlogram the existence of the phase angle ψ implies that the interval from $k = 0$ to the next maximum of the correlogram is not equal to the fundamental period. In judging the length of the period from the correlogram it is therefore better to measure from upcross to upcross or from trough to trough; or, if peaks are preferred, not to count the maximum at $k = 0$ as a peak.

The same result may be obtained in two other ways. Multiplying equation (2) by u_{t-k} and summing for all t we have
$$\rho_{k+2}+a\rho_{k+1}+b\rho_k = \frac{\Sigma(\epsilon_{t+2}u_{t-k})}{\operatorname{var} u}.$$

Since u_{t-k} depends only on ϵ_{t-k} and terms with lower subscripts, the expression on the right vanishes if k is not less than -1. We then have
$$\rho_{k+2}+a\rho_{k+1}+b\rho_k = 0 \quad (k \geqslant -1). \tag{14}$$

This result is due to Walker (1931). It was pointed out by Wold that if we multiply (2) by u_{t+k+2} and sum we get
$$\rho_k+a\rho_{k+1}+b\rho_{k+2} = \frac{\Sigma(\epsilon_{t+2}u_{t+k+2})}{\operatorname{var} u}.$$

The expression on the right no longer vanishes. In fact u_{t+k+2} contains the term $\xi_{k+1}\epsilon_{t+2}$ and we have
$$\rho_k+a\rho_{k+1}+b\rho_{k+2} = \frac{\operatorname{var}\epsilon}{\operatorname{var} u}\xi_{k+1} \quad (k \geqslant -1). \tag{15}$$

If $k = -1$ the two equations become identical, for $\rho_{-1} = \rho_1$ and we have
$$\rho_1(1+b)+a = 0, \quad \rho_1 = -\frac{a}{1+b}. \tag{16}$$

If we now solve either of the difference equations (14) and (15), making use of the initial conditions $\rho_0 = 1$ and (16) we arrive back at equation (13).

7. The foregoing result (13) would lead us to suppose that the correlogram of an autoregressive series would be damped according to the factor p^k, and this is true for an infinite series. In a number of practical cases, however, I was puzzled by the fact that correlograms of series which appeared on the face of it to be of type (2) did not damp out in the required way. Figs. 1 and 2 show the correlograms of two series for wheat prices and sheep population. The original series were taken from the agricultural returns for England and Wales, trend eliminated by a nine-years moving average and the first thirty serial correlations computed for the resulting series of 64 or 65 terms. The data and the correlations for what prices are given in Tables 1 and 2; those for sheep have been given in a previous paper (Kendall, 1941). I found a similar effect as regards non-damping in nine other agricultural time series, though the correlograms were not so regular as in these two cases. It was always possible, however, that the failure of the fluctuations to damp out in the correlogram was due wholly or partly

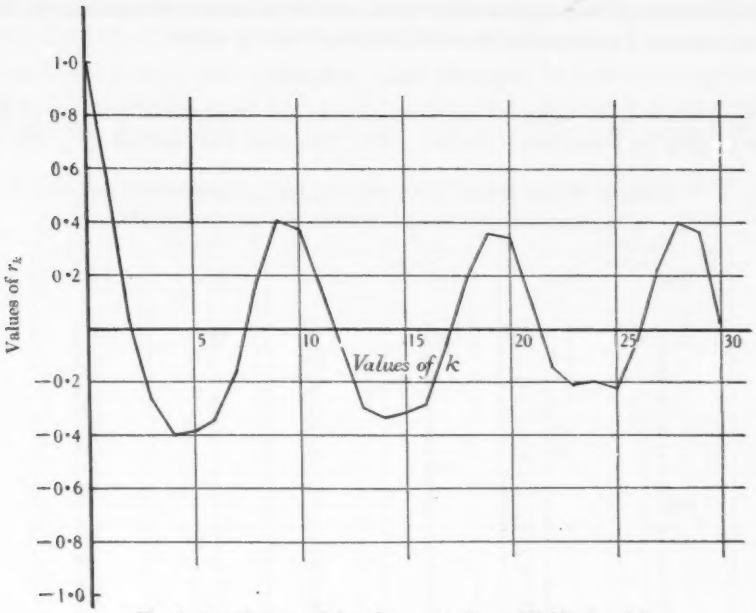


Fig. 1. Correlogram of the wheat price data of Tables 1 and 2.

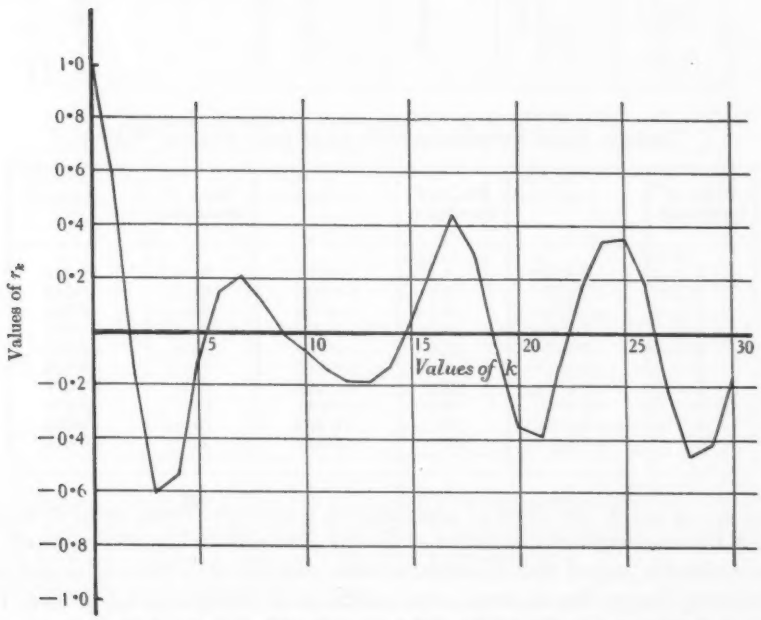


Fig. 2. Correlogram of the sheep population data referred to in paragraph 7.

to failure on the part of the original series to conform to the scheme of equation (2). To avoid such complications I constructed an artificial series with equation

$$u_{t+2} = 1{\cdot}1u_{t+1} - 0{\cdot}5u_t + \epsilon_{t+2}. \tag{17}$$

The ϵ's were taken from tables of random numbers and consisted of positive or negative numbers ranging by units from $-9{\cdot}5$ to $+9{\cdot}5$. The series was 'started up' from zero by

Table 1. *Wheat prices, 1871–1934 inclusive, deviations from nine-years moving average in pence per hundredweight*

Year	Price	Year	Price	Year	Price
1871	− 5	1893	6	1915	− 3
2	−13	4	18	6	− 7
3	−21	5	15	7	−40
4	− 6	6	1	8	−28
5	19	7	−11	9	−27
6	11	8	−17	1920	−51
7	−20	9	5	1	−14
8	3	1900	3	2	37
9	8	1	4	3	42
1880	2	2	2	4	11
1	− 6	3	3	5	− 9
2	−11	4	− 1	6	−17
3	− 7	5	2	7	−15
4	6	6	5	8	− 5
5	9	7	0	9	−10
6	9	8	− 3	1930	− 2
7	4	9	− 9	1	19
8	3	1910	1	2	14
9	4	1	9	3	21
1890	− 3	2	11	4	19
1	−20	3	32		
2	− 5	4	33		

Table 2. *Serial correlations of the wheat price series of Table* 1

Order of correlation	r	Order of correlation	r	Order of correlation	r
1	+0·577	11	+0·171	21	+0·115
2	+0·025	12	−0·075	22	−0·144
3	−0·267	13	−0·302	23	−0·204
4	−0·402	14	−0·332	24	−0·195
5	−0·389	15	−0·317	25	−0·221
6	−0·340	16	−0·282	26	−0·052
7	−0·174	17	−0·049	27	+0·217
8	+0·166	18	+0·200	28	+0·404
9	+0·411	19	+0·360	29	+0·364
10	+0·372	20	+0·343	30	+0·024

assuming $u_0 = u_{-1} = 0$. The effect of assuming two consecutive terms equal to zero is not important, for the contribution to a term of the series of terms far back is very small. We are therefore entitled to expect that an artificial series constructed in this way should conform to the foregoing theory. The resultant series and the serial correlations are given in Tables 3 and 4 and the correlogram in Fig. 3. Now this series is heavily damped, p being $\sqrt{(0{\cdot}5)} = 0{\cdot}7071$.

At the twentieth serial correlation, according to equation (14), r_{20} should be less than 0·002 in absolute magnitude. Actually it is one hundred times as big.

8. The explanation lies, I think, in the shortness of the series. In arriving at equation (13) it was assumed that product sums such as $\Sigma\epsilon_j\epsilon_{j+k}$ were zero; and this is so for a series of infinite length. But when the series is short these sums may differ quite appreciably from

Table 3. *Artificial series* $u_{t+2} = 1{\cdot}1u_{t+1} - 0{\cdot}5u_t + \epsilon_{t+2}$ *constructed as described in the text*

No. of term	Value of series	No. of term	Value of series	No. of term	Value of series
1	7	23	− 4	45	−13
2	6	24	− 5	46	1
3	− 6	25	− 9	47	6
4	− 4	26	− 4	48	4
5	3	27	− 4	49	11
6	− 4	28	3	50	15
7	− 5	29	9	51	9
8	− 1	30	4	52	8
9	10	31	− 8	53	4
10	10	32	− 6	54	− 1
11	6	33	− 3	55	4
12	− 4	34	− 2	56	7
13	− 4	35	0	57	11
14	− 7	36	− 1	58	0
15	− 2	37	− 3	59	1
16	6	38	3	60	0
17	17	39	− 1	61	− 5
18	24	40	− 8	62	−11
19	17	41	− 3	63	− 8
20	4	42	− 8	64	− 3
21	1	43	−10	65	5
22	− 5	44	−16		

Table 4. *Serial correlations of the series of Table* 3

Order of correlation	r	Order of correlation	r	Order of correlation	r
1	+0·70	11	−0·05	21	+0·05
2	+0·29	12	−0·17	22	−0·12
3	+0·01	13	−0·27	23	−0·28
4	−0·17	14	−0·31	24	−0·43
5	−0·27	15	−0·30	25	−0·57
6	−0·25	16	−0·18	26	−0·56
7	−0·13	17	+0·12	27	−0·26
8	+0·07	18	+0·29	28	+0·02
9	+0·12	19	+0·33	29	+0·17
10	+0·05	20	+0·22	30	+0·27

zero. For instance, in samples from a normal population with zero correlation, the standard error of the correlation in samples of 65 is about 0·125, so that values of 0·25 are not very improbable and even values as high as 0·375 are not impossible.

Consider then the series of n terms such as

$$u_j = \xi_1\epsilon_j + \xi_2\epsilon_{j-1} + \xi_3\epsilon_{j-2} + \ldots.$$

The product moment of u_j and u_{j+k} will be

$$\sum_j u_j u_{j+k} = \sum_j \sum_l \xi_l \epsilon_{j-l+1} \sum_m \xi_m \epsilon_{j+k-m+1} = \sum_j \sum_{l,m} \xi_l \xi_m \epsilon_{j-l+1} \epsilon_{j+k-m+1}.$$

The terms in ϵ^2 give the expression (13). The others will be sums of products of the ξ's multiplied by the serial covariances of the ϵ's themselves. The dominating terms in these latter will be those containing the larger ξ's. These terms are themselves damped harmonics of the fundamental type and when applied to the sums $\Sigma \epsilon_j \epsilon_{j+k}$ may be expected to generate oscillatory movements of about the same period as the original series. We may therefore expect that for short series the correlogram of the autoregressive system may not decay very rapidly, but that the product terms may themselves result in a small fluctuation. This appears to be happening in both the practical and the artificial examples given above.

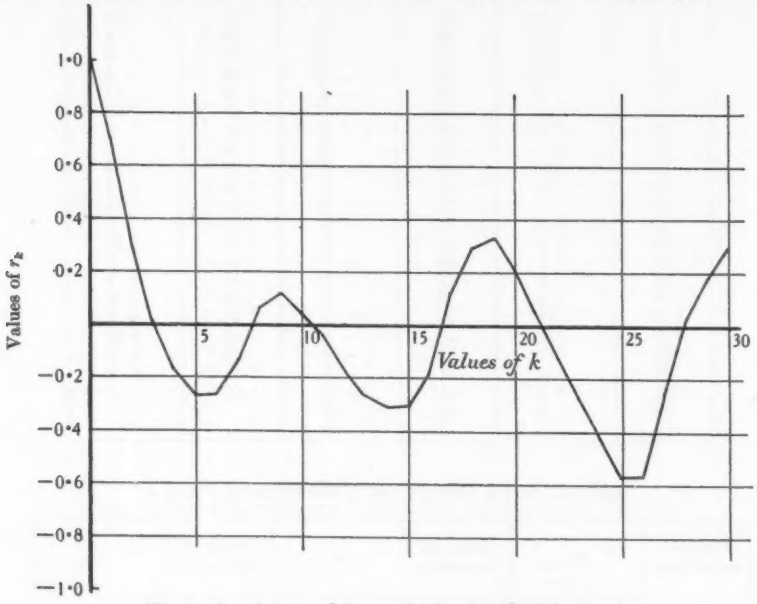


Fig. 3. Correlogram of the artificial series of Tables 3 and 4.

9. Putting aside the mathematics for a moment and looking at the point in a general way, we can, I think, appreciate that something of the kind is to be expected in short series. The variance of a short series should not differ systematically from the variance of the infinite series; but the covariances may systematically exceed in absolute value the value for the infinite series. In fact, as we proceed along the series, the oscillations change in phase, and when we have gone far enough will be quite unrelated in phase to the initial oscillations; but if we only go so far as the second or third oscillation the last oscillation may not, so to speak, have had time to get very much out of phase with the first. The consequence will be that the correlations for such a short space of time will tend to be higher than those for the series as a whole. For the generated system (and indeed for time series in general) we have to be careful not to assume too lightly that values calculated for a part of the series are typical of the corresponding values for the series as a whole; and this notwithstanding that the part of the series was chosen 'at random'.

One would like to know how long a practical series must be for the damping to show itself decisively. An exact discussion of this point presents difficulty, because in a finite series the serial covariances $\Sigma\epsilon_j\epsilon_{j+k}$ are not only non-vanishing but are correlated among themselves. Suppose, however, that we have a series with a damping factor p. Then the kth serial coefficient will not exceed p^k in absolute value, and the difference of the correlogram as a whole from expectation will not at any point exceed the error due to the sampling fluctuation of the serial correlation between u_j and u_{j+k}. This in turn will not exceed the sampling error of the correlation in samples of n from an uncorrelated series, which for large samples has a standard error $n^{-\frac{1}{2}}$. For example, if we took a series of 650 terms instead of 65, correlations up to the 30th would have a standard error not exceeding $1/\sqrt{(650-30)} = 0{\cdot}04$, and hence fluctuations of 0·10 would not be impossible. But by the 30th term in the correlogram the serial correlation has been practically damped out of existence. The fair inference is, I think, that except for the first few serial correlations of the main series, which are not damped very much, the serial correlations are seriously affected even for long series, for the sampling errors are not negligible compared with the small damped 'true' values. In general one would expect the damping effect to show itself for the first five, ten or twenty terms and then to be obliterated by the sampling effect. Whether this happens at the fifth, tenth or twentieth term depends not only on ϵ but on p, the rapidity of damping. In most of the natural series I have examined the damping is fairly rapid, so that the damping effect in the correlogram disappeared after the first few terms. The economic examples given by Wold (1938) and the meteorological examples of Walker (1931) appear to me to support this conclusion.

10. I turn now to consider another effect which may obscure the presence of the generated system of type (2) and may exert an important influence on the correlogram. The random element ϵ considered up to this point is what Yule calls a disturbance, and is integrated into the course of the series by the autoregressive scheme. There may also be a component η *superposed* on the system. This superposed element, if random, is like an error of observation in that its value at any point is unrelated to its value at other points.

If a random element with variance var η is superposed on an infinite series with variance var u the variance of the whole will be var η + var u. The autocovariances, however, will not be affected (except by sampling effects for short series). Consequently all the autocorrelations except ρ_0 will be reduced in the ratio

$$c = \frac{\text{var } u}{\text{var } u + \text{var } \eta}. \tag{18}$$

For short series there will still be a reduction of the same type, but the value of c may differ from its theoretical value for sampling reasons.

11. An autoregressive series of 65 terms was constructed according to the formula

$$u_{t+2} = 1{\cdot}5u_{t+1} - 0{\cdot}9u_t + \epsilon_{t+2}, \tag{19}$$

where the ϵ_{t+2} were the values of random numbers proceeding by units from $-49{\cdot}5$ to 49·5 with a theoretical variance of 833. On to the series so derived there were superposed (*a*) a rectangular random element $-49{\cdot}5(1)$ 49·5 and (*b*) a further rectangular random element $-199{\cdot}5(1)$ 199·5, additional to the first, the latter series then being divided by ten and rounded up to the nearest integer. The resultant series and serial correlations are shown in Tables 5 and 6, and the correlograms in Fig. 4.

According to (12) the variance of u for an infinite series generated by (19) should be 13·97 var ϵ. The values of c for the two series here considered are then, respectively,

$13{\cdot}97/14{\cdot}963 = 0{\cdot}93$ and $13{\cdot}97/30{\cdot}963 = 0{\cdot}45$. In the second case the effect of the superposed variation is to halve the observed correlations. The effect on the actual series of 65 terms is somewhat irregular (see para. 14 below).

Table 5. *Artificial series* $u_{t+2} = 1{\cdot}5u_{t+1} - 0{\cdot}9u_t + \epsilon_{t+2}$, *(a) with small superposed element* η, *(b) with large element* η, *constructed as described in the text*

No. of term	Value of series		No. of term	Value of series		No. of term	Value of series	
	(a)	(b)		(a)	(b)		(a)	(b)
1	5	16	23	−215	−34	45	88	3
2	36	7	24	−219	−15	46	76	− 5
3	8	10	25	− 76	−17	47	93	21
4	− 81	− 5	26	95	− 4	48	34	8
5	− 89	7	27	239	15	49	60	15
6	− 15	− 7	28	316	23	50	− 69	−13
7	35	10	29	289	21	51	−120	−10
8	112	13	30	169	13	52	− 54	12
9	146	17	31	49	− 7	53	− 56	−13
10	100	19	32	−114	−26	54	12	7
11	1	7	33	−259	−26	55	5	−11
12	−131	1	34	−290	−26	56	13	−18
13	−195	− 6	35	−208	−29	57	3	− 1
14	−259	−27	36	− 31	− 4	58	13	11
15	−258	−37	37	21	12	59	− 4	9
16	−118	−28	38	109	18	60	− 71	−13
17	32	6	39	26	19	61	− 94	− 7
18	110	− 3	40	33	− 5	62	15	19
19	245	6	41	22	0	63	45	− 3
20	166	24	42	− 30	−10	64	138	19
21	79	15	43	51	−15	65	115	0
22	−177	−34	44	49	17			

Table 6. *Serial correlations of series (a) and (b) of Table 5*

Order of correlation	r		Order of correlation	r		Order of correlation	r	
	(a)	(b)		(a)	(b)		(a)	(b)
1	+0·78	+0·49	11	+0·33	+0·16	21	+0·21	+0·05
2	+0·33	+0·13	12	−0·04	−0·16	22	−0·07	−0·28
3	−0·22	−0·13	13	−0·38	−0·37	23	−0·28	−0·33
4	−0·63	−0·42	14	−0·58	−0·50	24	−0·32	−0·29
5	−0·75	−0·46	15	−0·52	−0·36	25	−0·22	−0·20
6	−0·59	−0·39	16	−0·23	−0·14	26	−0·01	−0·07
7	−0·22	−0·01	17	+0·14	+0·19	27	+0·16	+0·11
8	+0·19	+0·28	18	+0·43	+0·42	28	+0·23	+0·20
9	+0·48	+0·38	19	+0·55	+0·41	29	+0·13	+0·01
10	+0·53	+0·52	20	+0·45	+0·37	30	−0·11	−0·11

The correlograms run according to expectation. The effect of the bigger random element is to reduce the amplitude at the beginning of the series and to introduce some minor irregularities in the data, but not to affect substantially the lengths of the correlogram oscillations.

12. But here arises one important difficulty. Suppose we are given such series as these and require to estimate a and b, the constants of the generating equation. The procedure adopted by Yule was as follows: we take the observed serial correlations r_1 and r_2 as estimates of the autocorrelations. We then find the regression equation of u_{t+2} on u_{t+1} and u_t by the usual methods, assuming that the variance of the series is the same as the variance for the series less its first term and that the serial correlation r_1 for the whole series is the same as that for the series less its first term. The regression equation is

$$u_{t+2} = \frac{r_1(1-r_2)}{1-r_1^2}u_{t+1} + \frac{r_2-r_1^2}{1-r_1^2}u_t. \tag{20}$$

The observed regression equation is then taken as an estimate of the generating equation so that we have as estimates of a and b

$$-a = \frac{r_1(1-r_2)}{1-r_1^2}, \quad -b = \frac{r_2-r_1^2}{1-r_1^2} = \frac{r_2-1}{1-r_1^2} + 1. \tag{21}$$

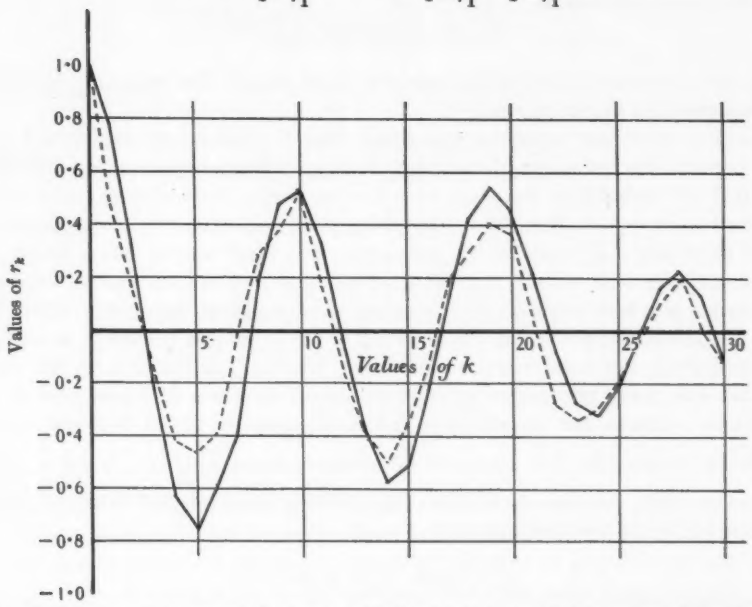


Fig. 4. Correlogram of the two artificial series of Tables 5 and 6, the full line representing series (*a*) with slight superposed variation, the broken line series (*b*) with large superposed variation.

From the above it will be evident that there are two sources of possible error in the use of these equations: (1) if the series is short r_1 and r_2 may not be reliable estimates of the autocorrelations for the infinite series, (2) if there is any superposed variation the observed r's will be lower than the true r's for the autoregressive system, being in fact cr_1 and cr_2 where c is given by (18). Consider the second of these effects.

13. In the case of superposed variation the use of (21) will lead to the equations

$$-a' = \frac{cr_1(1-cr_2)}{1-c^2r_1^2}, \quad -b' = \frac{cr_2-c^2r_1^2}{1-c^2r_1^2}.$$

The estimated fundamental period of the generating equation is then given by

$$4\cos^2\theta' = \frac{a'^2}{b'} = \frac{cr_1^2(1-cr_2)^2}{(1-c^2r_1^2)(r_2-cr_1^2)}. \tag{22}$$

If we expand (22) in powers of $\gamma = 1-c$ we find, to first order in γ,

$$\frac{a'^2}{b'} = \frac{a^2}{b}\left\{1-\gamma\frac{(1+b)(3b^2-b-a^2)}{b\{(1+b)^2-a^2\}}\right\}.$$

Hence, if $3b^2-b-a^2>0$ the effect of a superposed variation (equivalent to positive γ) is to give

$$\frac{a'^2}{b'} < \frac{a^2}{b},$$

or in other words to result in a shortening of the observed period. The condition that $3b^2-b-a^2>0$ is equivalent to

$$b > \tfrac{1}{6}\{-1+\sqrt{(12a^2+1)}\}$$

and is not very restrictive since in any case $a^2 \leqslant 4$ and $4b \geqslant a^2$. The inequality is obeyed by all the examples I have met in practice.

We therefore reach the interesting conclusion that if there is any superposed random variation present, the period calculated from the observed regression equation according to formulae (21) will probably be too short even for long series. Yule himself found too short a period for his sunspot material and, suspecting that it was due to superposed variation, attempted to reduce that variation by graduation. The result was to give a longer period more in accordance with observation. It does not appear, however, that the superposed variation in his case was very big. In a number of agricultural time series which I have examined it is sometimes about half the variation of the series and the effect on the period as calculated from the serial correlations is very serious. For instance, in the cases of wheat prices and sheep population referred to above, formulae (21) give periods of 7·0 and 6·8 years, whereas the correlograms indicate periods of about 9·5 and 8·5 years respectively.

14. As an example, consider the second of the artificial series referred to in paragraph 11. For the observed serial correlations I find

$$r_1' = 0{\cdot}486, \quad r_2' = 0{\cdot}133,$$

giving, according to (21),

$$-a' = 0{\cdot}552, \quad b' = 0{\cdot}135, \quad \cos\theta' = \frac{-a'}{2\sqrt{b'}} = 0{\cdot}751, \quad \theta' = 41{\cdot}3°,$$

which corresponds to a period of about 8·7 years, whereas we know from the construction of the series that

$$a = 1{\cdot}5, \quad b = 0{\cdot}9,$$

giving $\theta = 37{\cdot}7°$ and a period of 9·5 years.

Considering the profound effect which the superposed variation has had on the first two serial coefficients, reducing r_1 from 0·78 to 0·49 and r_2 from 0·33 to 0·13, one might have expected the period to have been affected even more than appears from this result. But the example serves to bring out once again the difficulties associated with short series and the unreliability of coefficients calculated from the first two serial correlations in such cases.

If, for instance, we had found $r_2' = 0{\cdot}18$ instead of $0{\cdot}13$ we should have obtained a period of about 12 years, and if $r_2' = 0{\cdot}20$ the solution becomes impossible, for a' and b' then assume values such that $a'^2 > 4b'$ and $\cos\theta' > 1$. Furthermore, such a value as $0{\cdot}18$ for r_2' increases the period instead of decreasing it. We note, in fact, that r_2 has been reduced by $0{\cdot}13/0{\cdot}33 = 0{\cdot}40$ so that it is not legitimate to assume that r_1 and r_2 are reduced by a constant c. The errors introduced by neglect of the autocovariances of the superposed random element may be so serious as to destroy the value of calculations based on the observed regression coefficients.

15. Even in long series, when it is legitimate to suppose that r_1 and r_2 are reduced in the same proportion, the length of the period is very sensitive to superposed variation. Consider, for example, the effect of small variations in c near $c = 1$. We have from (22), differentiating logarithmically and putting $c = 1$,

$$-2\tan\theta\frac{d\theta}{dc} = 1 - \frac{2r_2}{1-r_2} + \frac{2r_1^2}{1-r_1^2} + \frac{r_1^2}{r_2-r_1^2},$$

which on substituting $$r_1 = -\frac{a}{1+b}, \quad r_2 = -b + \frac{a^2}{1+b}$$

reduces to $$-\tan\theta\frac{d\theta}{dc} = \frac{(1+b)(3b^2+b-a^2)}{2b\{(1+b)^2-a^2\}}. \tag{23}$$

Now the period $$P = 2\pi/\theta \quad \text{and} \quad \tan\theta = \sqrt{\left(\frac{4b}{a^2}-1\right)}.$$

Hence $$\frac{dP}{dc} = -\frac{P^2a(1+b)(3b^2+b-a^2)}{4\pi\sqrt{(4b-a^2)}\{(1+b)^2-a^2\}}. \tag{24}$$

When $a = -1{\cdot}5$, $b = 0{\cdot}9$, $P = 9{\cdot}5$, this reduces to

$$\frac{dP}{dc} = 14 \text{ intervals.}$$

Thus if $c = 0{\cdot}9$, i.e. the superposed variation is only about 10 % of the total, the period may be shortened by something of the order of one interval.

16. The position may then be summarized as follows:

(*a*) The correlogram of a generated system of the type of equation (2) will be damped according to the damping factor of the equation; but if the series is short the damping may be considerably less than the theoretical value.

(*b*) The correlogram will show a period equal, within limits of error, to that of the fundamental period of the system. The distance from the unit ordinate at $k = 0$ and the first maximum of the correlogram may not, however, be a full period and in estimating periods from the correlogram it is better to reckon from upcross to upcross.

(*c*) It does not appear that in short series the periodic movement is substantially affected, at any rate not to such an extent as the damping.

(*d*) When superposed random variation is present the fundamental period calculated from the observed regressions will be too short and may be very considerably so.

(*e*) The period of the generated series, defined as the mean distance from upcross to upcross, is not quite the same as the fundamental period: but the difference is not likely to be important in practice.

17. Armed with these results we may consider the difficult problem of determining, for a given time series, the autoregressive scheme which may have generated it.

The first step is to calculate the serial correlations and examine the correlogram. If the latter shows fairly regular oscillations there is a presumption that the series is of the autoregressive type, and this conclusion need not be rejected solely because the oscillations do not damp so rapidly in the later portion of the correlogram as at the beginning. To take the wheat price material given above, an examination of the correlogram (Fig. 1) strongly suggests a simple damped harmonic. The sheep series (Fig. 2) is not so clearly defined and there is a suggestion of more than one period in the behaviour of the correlogram. This may, however, be due to the shortness of the series and one would be inclined to consider it in the first place as a single damped harmonic.

The length of the period, as pointed out above, is best determined by the mean distances between upcrosses. In the wheat price data there are upcrosses at about 7·5 years, 17·2 years and 26·1 years, giving periods of 9·7 and 8·9 years with a mean of 9·3 years. Much the same result is arrived at by counting the periods between troughs.

18. The second step is to calculate a' and b' by equations (21) and to find the period of the fundamental harmonic as given by the observed regression equation. For wheat prices we have

$$r'_1 = 0{\cdot}5773, \quad r'_2 = 0{\cdot}0246.$$

Hence

$$a' = -0{\cdot}8446, \quad b' = 0{\cdot}4630,$$

whence

$$\cos\theta' = 0{\cdot}6206, \quad \theta' = 51{\cdot}63°,$$

giving a fundamental period of 6·97 years.

This is too small, and we are immediately led to suspect the existence of superposed variation. The problem then arises of determining the variance of the superposed element. If this is random, and there are no periodic terms of very short period the variate difference method may be used. For the wheat price material, taking differences up to the 10th on the primary series (before trend was eliminated) I find for an estimate of the random variance $\operatorname{var}\eta = 27{\cdot}72$. The total variance of the series is 272·8. The constant c of equation (18) is

$$1 - \frac{27{\cdot}72}{272{\cdot}8} = 0{\cdot}90.$$

The putative serial correlations of the autoregressive series will then be

$$r_1 = 0{\cdot}641, \quad r_2 = 0{\cdot}027$$

and

$$a = -1{\cdot}059, \quad b = 0{\cdot}652,$$

whence

$$\cos\theta = 0{\cdot}6551, \quad \theta = 49{\cdot}07°,$$

giving a period of 7·34 years.

This is still too short. It would require a random superposed variance of about 25 % of the total, instead of the observed 10 %, to produce a period of 9·3 years.

19. This illustrates very well a constantly recurring difficulty in the theory of the variate difference method and of time series generally. The superposed variation may not be random. Indeed, we have little ground for expecting that it should be. A positive correlation between the successive values of η will reduce the variance shown as random by the variate difference method and unless we have prior reason to suppose that η is random the values given by the variate difference method are quite likely to be too small. Unfortunately we rarely have any prior knowledge of η, but from general economic considerations one would not be surprised

to find that there do exist positive correlations from one year to the next, owing to the enduring nature of some of the causes which can give rise to superposed variation. I conclude generally that discrepancies of the type here considered support the view that the period is to be determined from the correlogram, not from solution of the regression equation.

20. Two points may be mentioned incidentally. One, a matter of technique, is that the arithmetic of serial correlations and variate differences are closely linked together and the results of the one can be used to derive those of the other. This means an enormous saving in arithmetic and the method is described in the Appendix to this paper.

The second point concerns the removal of superposed random variation by graduation formulae. It will be clear that if the superposed variation is not random graduation may only make matters worse; but even if it is random, graduation formulae may induce spurious cyclical effects into the data. It seems to me that as a general rule graduation is to be undertaken with great caution.

21. Reverting to the main topic, it would seem that if the period shown by the correlogram and the period calculated from the observed regression equation disagree, and cannot be reconciled by the assumption of a superposed random element, there is little further to be done to dissect the superposed element from the autoregressive part of the system. If, however, the variate difference method supplies a variance of η which can satisfactorily explain the difference in periods, we may go forward. The constant c can be calculated and the constants of the autoregressive part of the system determined. Equation (12) then gives an estimate of the ratio between the variance of the basic random element ϵ and the variance of the generated series. If there is a superposed element it will not be possible to find the values of that element at every point and so to determine ϵ at every point; but if η is non-existent, ϵ can be explicitly determined, except for the first two terms of the series. In fact, we merely apply the generating equation to the observed series, the residuals between prediction and observation being the values of ϵ. An examination of these residuals will confirm whether they may be regarded as a random series.

22. The foregoing treatment can be extended to the case of a more general linear regressive system

$$u_{t+m} + a_1 u_{t+m-1} + \ldots + a_m u_t = \epsilon_{t+m},$$

or to cases where the regression is curvilinear; and the necessity for more general schemes in representing observed data can, as shown by Yule, be discussed in terms of partial autocorrelations and scatter diagrams. A more serious problem arises if the series ϵ is itself not random, a state of affairs which one fears might be fairly common in economic series. To take the wheat price data once again, it would not be surprising to find that the wheat price oscillations were regenerated by a series of disturbances, part of which were attributable to variations in acreages, yields, or the prices of other crops. Such disturbances might themselves be oscillatory. For such cases the problem becomes exceedingly complicated. To discuss it at all satisfactorily one would require a long series or collateral evidence in the form of other series of a similar character. If there is a royal road in this subject it has not yet been discovered.

APPENDIX

Relationship between variate differences and serial correlations

If we have a series of values $x_1 \dots x_n$ the first differences are $x_1 - x_2$, etc., the second differences $x_1 - 2x_2 + x_3$, etc., and so on. Let S_j be the sum of the squares of the jth differences and write for the product-sums

$$P_j = \sum_{k=1}^{n-j} x_k x_{k+j}. \tag{25}$$

The sums S are those appearing in variate difference analysis and the quantities P appear in serial correlation analysis. Either set can be expressed in terms of members of the other, as follows:

We have

$$S_0 = P_0, \tag{26}$$

$$S_1 = \Sigma (x_j - x_{j+1})^2 = \sum_{j=1}^{n-1} x_j^2 - 2\sum_{j=1}^{n-1} x_j x_{j+1} + \sum_{j=2}^{n} x_j^2 = 2P_0 - x_1^2 - x_n^2 - 2P_1, \tag{27}$$

$$S_2 = \Sigma (x_j - 2x_{j+1} + x_{j+2})^2 = \sum_{j=1}^{n-2} x_j^2 + 4\sum_{j=2}^{n-1} x_j^2 + \sum_{j=3}^{n} x_j^2 - 4\sum_{j=1}^{n-2} x_j x_{j+1} + 2\sum_{j=1}^{n-2} x_j x_{j+2} - 4\sum_{j=2}^{n-1} x_j x_{j+1}$$

$$= 6P_0 - 8P_1 + 2P_2 - x_1^2 - x_n^2 - (2x_1 - x_2)^2 - (2x_n - x_{n-1})^2 \tag{28}$$

and so on. For the purpose of expressing the general formulae of this kind it is convenient to modify the sums S. Suppose we write the series $x_1 \dots x_n$ preceded and followed by a number of zeros. The difference table will then appear as follows:

0				
	0			
0		0		
	0		$-x_1$	
0		$+x_1$		
	$-x_1$		$3x_1 - x_2$	etc.
x_1		$-2x_1 + x_2$		
	$x_1 - x_2$		$-3x_1 + 3x_2 - x_3$	
x_2		$x_1 - 2x_2 + x_3$		
	$x_2 - x_3$		$x_1 - 3x_2 + 3x_3 - x_4$	
x_3		$x_2 - 2x_3 + x_4$		

with a symmetrical effect at the other end. Writing now T_1, T_2, etc., for the sum of the squares of members in the first, second, etc., column of differences we see that

$$T_j = \sum_k \left\{ x_k - \binom{j}{1} x_{k+1} + \binom{j}{2} x_{k+2} \cdots \right\}^2, \tag{29}$$

where the summation now takes place over all values of x and there are no complications introduced by end effects. In fact, we have thrown the end effects into the sums T which replace the S's. In actually calculating the T's from the S's it is very little trouble to add the extra terms to the tables giving the latter; and when calculating the S's from the T's only the differences at the end of the table need be worked out.

We have then from (29), on expansion

$$T_j = P_0\left\{\binom{j}{0}^2 + \binom{j}{1}^2 + \ldots\right\} + 2(-1)^k \sum_k \left[P_k\left\{\binom{j}{0}\binom{j}{k} + \ldots + \binom{j}{j-k}\binom{j}{j}\right\}\right]. \tag{30}$$

The coefficients of the various P's are easily seen to be equal to corresponding powers of t in

$$\left\{\binom{j}{0}t^0 - \binom{j}{1}t + \binom{j}{2}t^2 \ldots\right\}\left\{\binom{j}{0}t^j - \binom{j}{1}t^{j-1} + \ldots\right\},$$

i.e. in $(-1)^j(1-t)^{2j}$, and we find, on substitution in (30),

$$T_j = P_0\binom{2j}{j} - 2P_1\binom{2j}{j-1} + 2P_2\binom{2j}{j-2} + \ldots + 2(-1)^j P_j. \tag{31}$$

For example

$$\begin{aligned}
T_0 &= P_0,\\
T_1 &= 2P_0 - 2P_1,\\
T_2 &= 6P_0 - 8P_1 + 2P_2,\\
T_3 &= 20P_0 - 30P_1 + 12P_2 - 2P_3,\\
T_4 &= 70P_0 - 112P_1 + 56P_2 - 16P_3 + 2P_4,\\
T_5 &= 252P_0 - 420P_1 + 240P_2 - 90P_3 + 20P_4 - 2P_5,\\
T_6 &= 924P_0 - 1584P_1 + 990P_2 - 440P_3 + 132P_4 - 24P_5 + 2P_6,\\
T_7 &= 3432P_0 - 6006P_1 + 4004P_2 - 2002P_3 + 728P_4 - 182P_5 + 28P_6 - 2P_7,\\
T_8 &= 12870P_0 - 22880P_1 + 16016P_2 - 8736P_3 + 3640P_4 - 1120P_5 + 240P_6 - 32P_7 + 2P_8,\\
T_9 &= 48620P_0 - 87516P_1 + 63648P_2 - 37128P_3 + 17136P_4 - 6120P_5 + 1632P_6 - 306P_7\\
&\quad + 36P_8 - 2P_9,\\
T_{10} &= 184756P_0 - 335920P_1 + 251940P_2 - 155040P_3 + 77520P_4 - 31008P_5 + 9690P_6\\
&\quad - 2280P_7 + 380P_8 - 40P_9 + 2P_{10}.
\end{aligned} \tag{32}$$

The coefficients check in virtue of the fact that they sum to zero.

Conversely we have

$$\begin{aligned}
2P_0 &= 2T_0,\\
2P_1 &= -T_1 + 2T_0,\\
2P_2 &= T_2 - 4T_1 + 2T_0,\\
2P_3 &= -T_3 + 6T_2 - 9T_1 + 2T_0,\\
2P_4 &= T_4 - 8T_3 + 20T_2 - 16T_1 + 2T_0,\\
2P_5 &= -T_5 + 10T_4 - 35T_3 + 50T_2 - 25T_1 + 2T_0,\\
2P_6 &= T_6 - 12T_5 + 54T_4 - 112T_3 + 105T_2 - 36T_1 + 2T_0,\\
2P_7 &= -T_7 + 14T_6 - 77T_5 + 210T_4 - 294T_3 + 196T_2 - 49T_1 + 2T_0,\\
2P_8 &= T_8 - 16T_7 + 104T_6 - 352T_5 + 660T_4 - 672T_3 + 336T_2 - 64T_1 + 2T_0,\\
2P_9 &= -T_9 + 18T_8 - 135T_7 + 546T_6 - 1287T_5 + 1782T_4 - 1386T_3 + 540T_2 - 81T_1 + 2T_0,\\
2P_{10} &= T_{10} - 20T_9 + 170T_8 - 800T_7 + 2275T_6 - 4004T_5 + 4290T_4 - 2640T_3 + 825T_2\\
&\quad - 100T_1 + 2T_0.
\end{aligned} \tag{33}$$

The coefficients sum in turn to 2, 1, −1, −2, −1, 1, 2, etc. The coefficient of T_k in $2P_j$ is

$$(-1)^{j-k}\frac{2j}{j+k}\binom{j+k}{2k}.$$

REFERENCES

DAVIS, H. T. (1941). *The Analysis of Economic Time Series.* Bloomington, Indiana.

DODD, E. L. (1939). The length of the cycles which result from the graduation of chance elements. *Ann. Math. Statist.* **10**, 254.

KENDALL, M. G. (1941). The effect of the elimination of trend on oscillations in time series. *J. Roy. Statist. Soc.* **104**, 43.

YULE, G. UDNY (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philos. Trans.* A, **226**, 267.

WALKER, SIR GILBERT (1931). On periodicity in series of related terms. *Proc. Roy. Soc.* A, **131**, 518.

WOLD, H. (1938). *A Study in the Analysis of Stationary Time-series.* Uppsala.

COMPARISON OF THE CONCEPTS OF EFFICIENCY AND CLOSENESS FOR CONSISTENT ESTIMATES OF A PARAMETER

By R. C. GEARY

Having given two estimates X and Y of an unknown parameter θ, E. J. G. Pitman (1937) has suggested that X should be regarded as a 'closer' estimate of θ if the probability that

$$|X-\theta|<|Y-\theta| \tag{1}$$

is greater than $\frac{1}{2}$. This concept has intuitive appeal and it is in accordance with statistical tradition that preference should be expressed on a probability scale. The object of this communication is to compare 'closeness', as defined, with the familiar concept of relative 'efficiency' (Fisher, 1922) as determined by the variances of the two estimates. Continuous variation is assumed throughout.

Pitman has shown that the median has a special role in his theory of closeness and, since the median is notably unamenable to algebraic treatment, it is not to be expected that, despite its apparent simplicity, condition (1) should be readily expressible in terms of the semi-invariants (assuming these known firmly or approximately) of the joint distribution of X and Y.

Suppose that X and Y are consistent estimates of θ, i.e. that

$$EX = EY = \theta. \tag{2}$$

In this connexion it must be observed that, in not necessarily large sample theory, this condition need not be observed. In fact, Pitman (1937, p. 215) has shown that in estimating the variance of normal samples of n, the 'closest' estimate of the variance is

$$s_2' = \sum_{i=1}^{n}(x_i-\bar{x})^2/(n-\tfrac{2}{3}), \text{ approximately,} \tag{3}$$

which is not consistent, instead of the usual

$$s_2'' = \sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1), \tag{4}$$

which is.

If the joint distribution of X and Y is symmetrical, i.e. if

$$\lambda_{ij} = \lambda_{ji} \tag{5}$$

for all joint semi-invariants of powers i and j, then it is clear that

$$\text{Prob}\{|X-\theta|<|Y-\theta|\} = \text{Prob}\{|Y-\theta|<|X-\theta|\},$$

and, since the sum of the two probabilities is unity, each must be equal to $\frac{1}{2}$. It follows that, when the two estimates are distributed on the normal surface of error with equal variances, the probability of (1) is $\frac{1}{2}$, since in this case $\lambda_{ij} = \lambda_{ji}$, the semi-invariants being zero for all values of i and j except those for which $(i+j) = 2$, where the sets for (i, j) of $(2, 0)$ and $(0, 2)$

are the variances of X and Y respectively. It can readily be shown that the probability of (1) in the normal case, with variances of σ_x^2 and σ_y^2 and coefficient of correlation ρ is

$$\frac{1}{\pi}\tan^{-1}\left(\frac{2\sigma_x\sigma_y\sqrt{(1-\rho^2)}}{\sigma_x^2-\sigma_y^2}\right), \tag{6}$$

which shows that in this case the criteria of closeness and efficiency are identical, since, for all values of ρ, the probability is greater than, or less than, $\frac{1}{2}$ according as σ_x is respectively less than, or greater than, σ_y.

Under general conditions the joint distribution of

$$x = X-\theta \quad \text{and} \quad y = Y-\theta$$

will be of the form

$$f(x,y) = \sum_{i+j=3}^{\infty}\sum \frac{(-)^{i+j}}{i!\,j!}\lambda_{ij}\left(\frac{\partial}{\partial x}\right)^i\left(\frac{\partial}{\partial y}\right)^j \Phi(x,y)\,dx\,dy, \tag{7}$$

$\Phi(x,y)$ being the normal function

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2}-\frac{2\rho xy}{\sigma_x\sigma_y}+\frac{y^2}{\sigma_y^2}\right)\right\}\Big/\,2\pi\sqrt{(1-\rho^2)}. \tag{8}$$

If the two estimates are computed from samples of n it is well known that

$$\lambda_{ij}/\sigma_x^i\sigma_y^j$$

is of order $\frac{1}{2}(2-i-j)$ in n. It is natural, accordingly, to try to find the value of probability of (1) when the variates are distributed in (7) assuming that the λ_{ij} are zero for all but the smaller values of $(i+j)$. Probability of (1) is given by

$$\int_{-\infty}^{+\infty} dy \int_{-|y|}^{+|y|} f(x,y)\,dx. \tag{9}$$

To find the value of this integral the frequency $f(x,y)$ is regarded as the product of the normal function Φ and a polynomial in x and y determined from (7). It can easily be shown that polynomials of odd order in $(i+j)$ vanish on integration. Accordingly the integral (9) is given by (6) plus a term $O(n^{-1})$; furthermore the expression for the probability contains only terms in n^{-k}, where k is a positive integer.

There is no theoretical difficulty about the computation of the coefficients of powers of $\lambda_{ij}/\sigma_i\sigma_j$ in (9) but the algebra is complicated. For the purpose of this note it will suffice perhaps to state that when

$$\sigma_x = \sigma_y,\ \lambda_{ij} = 0 \text{ for } (i+j)>4 \quad \text{and} \quad \lambda_{ij}^k = 0 \text{ for } k>2,\ (i+j)=3,$$

then the probability (1) is given by

$$\begin{aligned}\frac{1}{2}+\frac{1}{72\pi}(1-\rho^2)^{-\frac{1}{2}}\{&6(1-\rho^2)(\lambda'_{40}-\lambda'_{04})-12\rho(1-\rho^2)(\lambda'_{31}-\lambda'_{13})\\ &-(7+2\rho^2)(\lambda'^2_{30}-\lambda'^2_{03})+36\rho(\lambda'_{30}\lambda'_{21}-\lambda'_{12}\lambda'_{03})\\ &-3(1+2\rho^2)(2\lambda'_{30}\lambda'_{12}+3\lambda'^2_{21}-2\lambda'_{21}\lambda'_{03}-3\lambda'^2_{12})\},\end{aligned} \tag{10}$$

with $$\lambda'_{ij} = \lambda_{ij}/\sigma_x^i\sigma_y^j.$$

The term in addition to $\frac{1}{2}$ may be regarded as $O(n^{-1})$, the terms neglected being $O(n^{-2})$ when n is large. When n is not large it is evident that the added term may be quite appreciable

when λ'_{ij} is significantly different from λ_{ji} and more particularly if ρ is nearly unity, which will usually be the case with two estimates which are of almost equal efficiency. In a particular example (which, however, had no reference to the problem of estimation) the values found were as follows:

$$\rho = 0{\cdot}953; \quad \lambda'_{30} = -0{\cdot}273; \quad \lambda'_{21} = -0{\cdot}287; \quad \lambda'_{12} = -0{\cdot}302; \quad \lambda'_{03} = -0{\cdot}307;$$

$$\lambda'_{40} = -0{\cdot}563; \quad \lambda'_{31} = -0{\cdot}513; \quad \lambda'_{22} = -0{\cdot}466; \quad \lambda'_{13} = -0{\cdot}418; \quad \lambda'_{04} = -0{\cdot}327.$$

For these values of the semi-invariants the value of (10) is 0·5185, which does not differ much from the value $\frac{1}{2}$ which would be obtained on the assumption of equal variances and normal distribution of the estimates; it is equally certain that, when the estimates are computed from large random samples and when their joint distribution tends towards normality as the sample number tends towards infinity, the probability (1) will be exceedingly close to $\frac{1}{2}$; there is no good *theoretical* reason, however, for thinking that for estimates computed from samples which are not large, the value of (10) (or of the more exact value of the probability which would be found by taking into account further terms in the expansion of (7) in the compilation of (9)) would be close to $\frac{1}{2}$, in general.

SOME PARTICULAR CASES

As an application, consider the method of estimating the universal mean of a normal distribution, give a random sample of n, assuming that the universal variance is unity. The distribution is, accordingly,

$$\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(z-\theta)^2}\, dz.$$

Since the maximum likelihood estimate of θ is the mean

$$X = \frac{1}{n} \sum_{i=1}^{n} z_i$$

its variance in large samples must be less than that of any other estimate Y of the parameter (Fisher, 1922, 1935). It is well-known that the median y of large samples of n from a normal universe with mean zero and variance unity is distributed approximately as

$$\frac{1}{\pi} \sqrt{\left(\frac{n}{2}\right)} e^{-ny^2/\pi}\, dy,$$

whereas the mean x is distributed as

$$\sqrt{\left(\frac{n}{2\pi}\right)} e^{-nx^2/2}\, dx,$$

so that the respective variances are $\sqrt{(\pi/2n)}$ and $1/\sqrt{n}$. It has been computed that for large normal samples the correlation between the mean and the median is $\rho = \sqrt{(2/\pi)}$. From formula (6)

$$\text{Prob}\{|X-\theta| < |Y-\theta|\} = 0{\cdot}615.$$

Pitman (1937, p. 221) has shown that the closest estimate of the centre of a rectangular distribution is the mid-point (or the mean of the largest and smallest members of the sample). The rest of this section deals with comparisons, by means of the criteria of closeness and efficiency, of the mid-point and other consistent estimates of the centre of the rectangular distribution of which the range is known to be unity.

It is a well-known fact that by the variance test the mid-point is more efficient than the *arithmetic mean* as an estimate of the middle point of the range for a rectangular universe when the range is known. In fact the respective variances are as follows:

	Variance
Mid-point:	$1/\{2(n+1)(n+2)\}$
Mean:	$1/\{12n\}$,

so that the variance of the mid-point is actually of a lower order of magnitude in n than the variance of the mean. The respective distributions (when the centre is zero) are as follows:

Mid-point:
$$n(1-2|t|)^{n-1}\,dt \tag{11}$$

Mean (Gram-Charlier, correct to n^{-3}):

$$\frac{dy}{\sqrt{(2\pi)}}\left[1-\frac{1}{20n}\left(\frac{d}{dy}\right)^4+\frac{1}{n^2}\left\{\frac{1}{105}\left(\frac{d}{dy}\right)^6+\frac{1}{800}\left(\frac{d}{dy}\right)^8\right\}\right.$$
$$\left.-\frac{1}{n^3}\left\{\frac{3}{1{,}400}\left(\frac{d}{dy}\right)^8+\frac{1}{2{,}100}\left(\frac{d}{dy}\right)^{10}+\frac{1}{48{,}000}\left(\frac{d}{dy}\right)^{12}\right\}+\ldots\right]e^{-y^2/2}, \tag{12}$$

$y = x\sqrt{(12n)}$, x being the mean of sample of n.

The distributions of the rectangular mid-point and lowest point (used later) have been given by Neyman & Pearson (1928). The β_2 of the distribution of the mid-point is given by

$$\beta_2 = \frac{6(n+1)(n+2)}{(n+3)(n+4)}, \tag{13}$$

which tends towards 6 (instead of the normal value 3) when n tends towards infinity. The distribution of the mean tends rapidly towards normality when n tends towards infinity. Formula (6) cannot be used in this case. Actually it is found that, when n is large,

$$\text{Prob}\{|x|<|t|\} = \sqrt{\left(\frac{6}{n\pi}\right)} \doteq 1{\cdot}38n^{-\frac{1}{2}}. \tag{14}$$

It is interesting to compare this probability with the probability which would be found from formula (6), i.e. on the assumption that both estimates were normally distributed. The coefficient of correlation between mean and mid-point is

$$\rho = \sqrt{\left(\frac{6n}{(n+1)(n+2)}\right)} = \sqrt{(6)}\,n^{-\frac{1}{2}} \tag{15}$$

and the pseudo-probability required is approximately

$$\frac{2\sqrt{6}}{\pi}n^{-\frac{1}{2}} = 1{\cdot}56n^{-\frac{1}{2}},$$

which is very little different from the true probability at (14).

In the previous application one of the two estimates (namely the mean) was approximately normally distributed. The mid-point is now compared with another estimate, the variance of which is of the same order of magnitude (in n) as that of the mid-point, but the distribution of which does not tend towards normality with increasing n.

This other estimate is the lowest point in the sample less its mean for $\theta = 0$. This estimate is clearly consistent. Since both estimates are consistent their deviations may be regarded as measured from θ, the unknown centre of the rectangular distribution.

In the range $(-\frac{1}{2}, +\frac{1}{2})$ the distribution of the lowest point is given by $n(\frac{1}{2}-u)^{n-1}du$, the mean value of which is $-k = -\frac{1}{2}(n-1)/(n+1)$. The proposed estimate is $(u+k)$. If the largest value is v the mid-point estimate is $t = \frac{1}{2}(u+v)$. Hence it is required to consider the probability of

$$|u+k| < \tfrac{1}{2}|u+v|. \tag{16}$$

The joint distribution of u and v is given by

$$n(n-1)|v-u|^{n-2}du\,dv. \tag{17}$$

In the (u, v) plane the inequalities (10) define two 'critical' straight lines given by

$$v-u-2k=0 \quad \text{and} \quad v+3u+2k=0.$$

The required probability will be found by integrating (17) over certain areas bounded by these lines and by the lines $u = \pm\frac{1}{2}$, $v = \pm\frac{1}{2}$. The probability is a very complicated function of n which, however, reduces to

$$\tfrac{3}{4}e^{-1} - \tfrac{1}{2}e^{-2} = 0{\cdot}317, \tag{18}$$

when $n = \infty$. Accordingly the mid-point is a much closer estimate of θ than the lowest point less its mean for $\theta = 0$.

As regards efficiency, the variances of u and t and the coefficient of correlation between u and t are as follows:

$$\sigma_u^2 = n/\{(n+1)^2(n+2)\}, \quad \sigma_t^2 = 1/\{2(n+1)(n+2)\}, \quad \rho_{ut} = \sqrt{(n+1)}/\sqrt{(2n)},$$

so that, if u and t were normally distributed, the probability that $|u-\bar{u}| < |t|$, from (6), is given by

$$\frac{1}{\pi}\tan^{-1}(2\sqrt{(n+1)}/\sqrt{(n-1)})$$

which, when n tends towards infinity, tends towards 0·352. This is not very different from the true probability 0·317.

A FURTHER ILLUSTRATION: 'CARS IN A TOWN'

At a recent meeting of the Dublin University Mathematical Society, E. Schrödinger suggested the following ingenious problem as an illustration of Pitman's concept of *closeness*. In a town, cars are known to be numbered consecutively from 1. The numbers on r of the cars are noted: the problem is to find the closest estimate of the number of cars in the town. Following is the solution, on Pitman's lines, of Schrödinger's problem.

Let n, the unknown total number, be assumed to be so large* that variation is continuous, i.e. that any car number observed at random has a rectangular frequency distribution. The highest of the r numbers observed, namely w, is a sufficient statistic for n because, when w is known, the remaining $(r-1)$ variates have a joint frequency distribution independent of n; hence all relevant information can be derived from w; the remaining $(r-1)$ observations may be ignored. The cumulative frequency distribution of w is

$$w^r/n^r.$$

By Pitman's theory the closest estimate will be that for which the observed w has the median value, i.e. if $\hat{n}$ is the estimate of n,

$$(w/\hat{n})^r = \tfrac{1}{2} \quad \text{or} \quad \hat{n} = 2^{1/r}w.$$

It is clear that $2^{1/r}w$ has median value n. If $r = 1$, $\hat{n} = 2w$, and if r tends towards ∞, $\hat{n}$ tends towards w, both of which are reasonable. That $\hat{n}$ actually is the closest estimate transpir

* It seems likely that the solution is valid for all values of n.

from a theorem of Pitman (1937) to the effect that if $\hat{n}$ has median value n, and if n' be any other estimating function, then $\hat{n}$ is a closer estimate of n than n', i.e.

$$\text{Prob}\{|\hat{n}-n| < |n'-n|\} > \tfrac{1}{2},$$

provided that a variable z can be found, always of the same sign, so that

$$\hat{n} \quad \text{and} \quad z(n'-\hat{n})$$

are independent.

In the present application suppose that n' is another estimate of n. From considerations of scale, n' must be homogeneous and of degree 1 in the observation w and the other observations $w_1, w_2, \ldots, w_{r-1}$. Hence $n'/\hat{n}$ must be homogeneous and of degree 0 in $w, w_1, w_2, \ldots, w_{r-1}$, and therefore expressible as a function of

$$q_i = w_i/w, \quad i = 1, 2, \ldots, r-1.$$

Now it is obvious that the q_i, and hence $n'/\hat{n}$, are independent of w. Therefore taking z as $1/\hat{n}$, which is always positive, we see that $\hat{n}$ is the closest estimate.

To find an upper limit (in probability) of n, we express the fact that n should not be so great as to render too unlikely the occurrence of the largest number actually observed w. Accordingly, pre-determine a probability α and set

$$(w/n)^r \geqslant \alpha.$$

Hence

$$n \leqslant w/\alpha^{1/r}.$$

Example. The number of 30 motor cars are noted and the largest number is found to be 247: to estimate the number of cars in the town and the upper limit of error of the estimate. Here $r = 30$, $w = 247$, so that

$$n = 2^{1/30} \times 247 = 253 \text{ approximately,}$$

and

$$n \leqslant 247/(0{\cdot}05)^{1/30} = 273 \text{ approximately, if } \alpha = 0{\cdot}05.$$

The latter statement means that the number of cars in the town will be less than 273 unless in taking the particular sample an event, the probability of which was 1/20, occurred.

Summary and conclusion

The criterion of *efficiency* as determined by a comparison of the variances of two estimates of an unknown statistic is identical with the criterion of *closeness* when the joint distribution of the estimates is normal and the criteria will not yield significantly different results in practice when the estimates are estimated from large samples and are consistent. Study of some particular examples suggests that, even when the distribution of the estimates is very different from normal the value of the probability associated with the criterion of closeness may not be very different from what its value would be if normality of the joint distribution were assumed. An application of Pitman's theory of closeness is discussed.

The author is much indebted to Prof. E. S. Pearson for his helpful criticism of this Note, and to Mr N. L. Johnson for pointing out an error in algebra in the first draft submitted.

References

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philos. Trans.* **A, 122**, 309.

Fisher, R. A. (1935). The Logic of Inductive Inference. *J. Roy. Statist. Soc.* **98, 39.**

Neyman, J. & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20 A**, 175.

Pitman, E. J. G. (1937). The 'closest' estimates of statistical parameters. *Proc. Camb. Phil. Soc.* **33**, 212.

THE RELATION BETWEEN MEASURES OF CORRELATION IN THE UNIVERSE OF SAMPLE PERMUTATIONS

By H. E. DANIELS, *Wool Industries Research Association*

1. Introduction

Recent papers by Hotelling & Pabst (1936), Pitman (1937), Kendall (1938) and Kendall, Kendall & Babington-Smith (1939) discuss the distribution of the correlation coefficient when members of the sample corresponding to the two variates are permuted randomly relative to one another. In the case of rank correlation, the characteristics of the population sampled are generally unknown, and a significance test has to be based on the distribution obtained from the sample in this way.

Hotelling & Pabst prove that as the sample size is increased, Spearman's ρ tends to follow a normal distribution law. Kendall's measure of rank correlation, τ, in which all possible corresponding pairs in two given rankings are assigned marks ± 1 according to whether they agree or differ in order, follows a specially simple distribution law which tends rapidly to the normal form and becomes highly correlated with Spearman's ρ for samples of moderate size.

The present paper discusses the properties of the class of correlation coefficients Γ obtained on replacing Kendall's marks ± 1 by a more general system of scores. By an empirical argument Kendall *et al.* showed it to be likely that the correlation between τ and ρ is $\frac{2(n+1)}{\sqrt{[2n(2n+5)]}}$ for all values of the sample size n, and surmised that their joint distribution tends to the bivariate normal form for large n. These results are, in fact, special cases of the relations demonstrated below between two correlation coefficients Γ with different systems of scores.

2. Definition

Consider the two sets of n sample values

$$x_1, x_2, \ldots, x_n, \quad y_1, y_2, \ldots, y_n,$$

both arranged in some given order relative to each other. They may be permuted to give $n!$ different ways of grouping the x's with the y's. Let us assign to each pair (x_i, x_j) what for convenience will be termed a score a_{ij} and to each (y_i, y_j) a score b_{ij}, where

$$a_{ij} = -a_{ji}, \quad b_{ij} = -b_{ji}.$$

Denote by Γ the number

$$\Gamma = \frac{\Sigma a_{ij} b_{ij}}{\sqrt{(\Sigma a_{ij}^2 \, \Sigma b_{ij}^2)}},$$

the summation extending over all i and j from 1 to n. Special cases of Γ are Kendall's τ, the product-moment correlation coefficient r and Spearman's ρ, for τ is obtained by definition when $a_{ij}, b_{ij} = \pm 1, j \gtrless i$, r is given when $a_{ij} = x_j - x_i$, $b_{ij} = y_j - y_i$ by virtue of the identity

$$\tfrac{1}{2} \sum_i \sum_j (x_j - x_i)(y_j - y_i) \equiv n \sum_i x_i y_i - \sum_i \sum_j x_i y_j,$$

and ρ is similarly obtained when $a_{ij}, b_{ij} = j - i$.

When the x's are permuted relative to the y's, the scores reappear in a new order with the same or opposite sign and the denominator of Γ remains unaltered, so that in discussing

the distribution of Γ over all permutations it is sufficient to consider the numerator only, which we denote by c.

Write A, B for the matrices of the scores; for example, with $n = 4$,

$$A = \begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{bmatrix}; \quad B = \begin{bmatrix} 0 & b_{12} & b_{13} & b_{14} \\ -b_{12} & 0 & b_{23} & b_{24} \\ -b_{13} & -b_{23} & 0 & b_{34} \\ -b_{14} & -b_{24} & -b_{34} & 0 \end{bmatrix}.$$

With the x's and y's in the order as written, c is the trace of the matrix product AB' (i.e. the sum of the elements of its leading diagonal), where B' is the transpose of B. The effect of a permutation of the x's, say, is to alter the score matrix of the x pairs to PAP' and the value of c to the trace of $PAP'B'$, where $P = (p_{ij})$ is the appropriate 'permutation matrix' obtained by permuting the columns of the unit matrix. For example, corresponding to the grouping

$$x_4, x_1, x_2, x_3, \quad y_1, y_2, y_3, y_4,$$

the permutation matrix is

$$P = \begin{bmatrix} . & . & . & 1 \\ 1 & . & . & . \\ . & 1 & . & . \\ . & . & 1 & . \end{bmatrix}.$$

In terms of the matrix elements c is given by

$$c = \Sigma p_{ji} p_{kl} a_{il} b_{jk},$$

all suffixes being summed from 1 to n.

3. Moments

The distribution of c over all permutations, or the joint distribution of two c's with different systems of scores, is most readily discussed from its moments. The product moment of $c(1)$ and $c(2)$ with scores $a_{ij}^{(1)}$, $b_{ij}^{(1)}$ and $a_{ij}^{(2)}$, $b_{ij}^{(2)}$ respectively is

$$\overline{c(1)\,c(2)} = \frac{1}{n!} S(c(1)\,c(2)),$$

where

$$c(1)\,c(2) = \Sigma p_{ji} p_{kl} p_{sr} p_{tu} a_{il}^{(1)} b_{jk}^{(1)} a_{rt}^{(2)} b_{su}^{(2)},$$

and S denotes summation over all $n!$ possible permutations. Consider the effect of S on each term. The non-vanishing contributions occur when all four p's are 1, and it is first noted that if any of the suffixes i, l, r, t are equal, the corresponding suffixes in j, k, s, u must also be equal for the term not to vanish, since each row of P contains only one non-vanishing element. Terms in which $i = l$ or $r = t$ are of course zero by definition. When, for example, $i = r$ so that r is replaced by i in the expression, we shall call i a *tied suffix*. Other suffixes will be referred to as *free* suffixes.

As regards their contribution to S the terms may be classified according to the number of tied suffixes in i, l, r, t as follows.

(i) *No tied suffixes.* When the four p's are each unity, four rows and columns of P are assigned and there are $(n-4)!$ ways of filling the remaining positions. Such terms therefore contribute

$$(n-4)!\,\Sigma' a_{il}^{(1)} a_{rt}^{(1)} \Sigma' b_{jk}^{(1)} b_{su}^{(2)}$$

to the total sum, where Σ' denotes summation over all values of the suffixes which are not

equal. Let us consider some properties of $\Sigma' a_{ij}a_{rt}$ and $\Sigma a_{ij}a_{rt}$ and similar expressions with tied suffixes, the second expression being summed over *all* values of the suffixes. (Superscripts (1) and (2) are understood throughout.)

(1) $\Sigma a_{ii}a_{rt} = 0$, $\Sigma' a_{ii}a_{rt} = 0$, and so on.

(2) $\Sigma a_{ij}a_{ij} = \Sigma' a_{ij}a_{ij}$.

(3) $\Sigma a_{ij}a_{ij} = \Sigma' a_{ij}a_{ij} + \Sigma' a_{ij}a_{ij}$.

(4) $\Sigma a_{ij}a_{rt} = 0$, $\Sigma' a_{ij}a_{rt} = 0$. The first is true because $\Sigma a_{ij} = 0$, and the second follows from the fact that

$$\Sigma a_{ij}a_{rt} = \Sigma' a_{ij}a_{rt} + \Sigma' a_{ij}a_{ij} + \Sigma' a_{ij}a_{ri} + \Sigma' a_{ij}a_{ij} + \Sigma' a_{ij}a_{rj} + \Sigma' a_{ij}a_{ij} + \Sigma' a_{ij}a_{ii},$$

the terms on the right after the first cancelling in pairs.

The contribution to S of terms with no tied suffixes is therefore zero.

(ii) *One tied suffix.* For the term not to vanish it is necessary to assign three rows and columns of P, and the contribution to S from such terms is

$$4(n-3)!\,\Sigma' a_{ij}^{(1)} a_{ij}^{(2)} \Sigma' b_{jk}^{(1)} b_{ju}^{(2)},$$

the factor 4 arising from the fact that the same contribution is obtained by tying the suffixes in the four possible ways.

(iii) *Two tied suffixes.* The contribution to S is similarly found to be

$$2(n-2)!\,\Sigma' a_{ij}^{(1)} a_{ij}^{(2)} \Sigma' b_{jk}^{(1)} b_{jk}^{(2)}.$$

Terms containing more than two tied suffixes give zero contributions to S, and finally, substituting for Σ' the appropriate Σ expressions, we find

$$\overline{c(1)\,c(2)} = \frac{4}{n(n-1)(n-2)} (\Sigma a_{ij}^{(1)} a_{ij}^{(2)} - \Sigma a_{ij}^{(1)} a_{ij}^{(2)}) (\Sigma b_{jk}^{(1)} b_{ju}^{(2)} - \Sigma b_{jk}^{(1)} b_{jk}^{(2)}) + \frac{2}{n(n-1)} \Sigma a_{ij}^{(1)} a_{ij}^{(2)} \Sigma b_{jk}^{(1)} b_{jk}^{(2)}.$$

The moments of higher order can be obtained by a similar procedure, but the expressions rapidly become unwieldy.

4. The correlation between Kendall's τ and Spearman's ρ

As a first application of the formula we consider the correlation between τ and ρ over all permutations of the sample values. The scores for τ and ρ respectively are

$$a_{ij}^{(1)},\ b_{ij}^{(1)} = \pm 1,\ 0 \quad \text{when} \quad j \gtrless i,\ j = i,$$

$$a_{ij}^{(2)},\ b_{ij}^{(2)} = j - i.$$

The following results are easily derived

$$\sum_{l=1}^{n} a_{ij}^{(1)} = n + 1 - 2i, \quad \sum_{l=1}^{n} a_{ij}^{(2)} = \tfrac{1}{2} n(n+1-2i),$$

$$\sum_{i=1}^{n} \sum_{l=1}^{n} \sum_{t=1}^{n} a_{ij}^{(1)} a_{ij}^{(2)} = \frac{n^2(n^2-1)}{6}, \quad \sum_{i=1}^{n} \sum_{l=1}^{n} a_{ij}^{(1)} a_{ij}^{(2)} = \frac{n(n^2-1)}{3},$$

and the same results hold for the b's. Substitution in the formula then gives

$$\overline{c(1)\,c(2)} = \frac{n^2(n-1)(n+1)^2}{9}.$$

Again, $$\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{t=1}^{n}a_{il}^{(1)}a_{it}^{(1)}=\frac{n(n^2-1)}{3},\quad \sum_{i=1}^{n}\sum_{l=1}^{n}a_{il}^{(1)}a_{il}^{(1)}=n(n-1)$$

and $$\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{t=1}^{n}a_{il}^{(2)}a_{it}^{(2)}=\frac{n^3(n^2-1)}{12},\quad \sum_{i=1}^{n}\sum_{l=1}^{n}a_{il}^{(2)}a_{il}^{(2)}=\frac{n(n^2-1)}{3},$$

from which it is found that

$$\overline{c(1)^2}=\frac{2n(n-1)(2n+5)}{9},\quad \overline{c(2)^2}=\frac{n^4(n-1)(n+1)^2}{36}.$$

The required correlation is therefore

$$R_{\rho r}=\frac{2(n+1)}{\sqrt{[2n(2n+5)]}}$$

which is the result anticipated by Kendall *et al.* It should be noted that they use the quantities $\Sigma=\frac{1}{2}c(1)$ and

$$S(d^2)=\frac{n(n^2-1)}{6}-\frac{c(2)}{n}$$

in place of $c(1)$ and $c(2)$.

5. Transformation of the sample values

If the scales of the x's and y's are distorted by a transformation and the product moment correlation coefficient r is recalculated on the transformed sample, a new value r' is obtained. In particular, the x's and y's may be readjusted to be at equidistant intervals, and then the new value of r is Spearman's ρ. The formula for $\overline{c(1)\,c(2)}$ can be used to find the correlation over all sample permutations between the values of r on the same sample before and after such a transformation. Distinguishing by primes the sample values after transformation, the scores are

$$a_{ij}^{(1)}=x_j-x_i,\quad b_{ij}^{(1)}=y_j-y_i,$$

$$a_{ij}^{(2)}=x'_j-x'_i,\quad b_{ij}^{(2)}=y'_j-y'_i.$$

Then $$\sum_{l=1}^{n}a_{il}^{(1)}=n(\bar{x}-x_i),\quad \sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{t=1}^{n}a_{il}^{(1)}a_{it}^{(2)}=n^2\sum_{i=1}^{n}(x_i-\bar{x})(x'_i-\bar{x}'),$$

$$\sum_{i=1}^{n}\sum_{l=1}^{n}a_{il}^{(1)}a_{il}^{(2)}=\sum_{i=1}^{n}\sum_{l=1}^{n}(x_l-x_i)(x'_l-x'_i)=2n\Sigma(x_i-\bar{x})(x'_i-\bar{x}')$$

by the identity previously quoted. Using these and similar formulae we find

$$\overline{c(1)\,c(2)}=\frac{4n^2}{n-1}\Sigma(x-\bar{x})(x'-\bar{x}')\,\Sigma(y-\bar{y})(y'-\bar{y}'),$$

$$\overline{c(1)^2}=\frac{4n^2}{n-1}\Sigma(x-\bar{x})^2\,\Sigma(y-\bar{y})^2;\quad \overline{c(2)^2}=\frac{4n^2}{n-1}\Sigma(x'-\bar{x}')^2\,\Sigma(y'-\bar{y}')^2,$$

and hence the correlation between r and r' is

$$R_{rr'}=r_{xx'}r_{yy'},$$

where $r_{xx'}$ and $r_{yy'}$ are the correlation coefficients between old and new values of x and y respectively.

6. TENDENCY TO NORMAL FORM FOR LARGE n

It will now be shown for a large class of score systems a_{ij} that c, and hence Γ, tends with increasing n to be normally distributed, and moreover, that the joint distribution of any pair of such Γ's tends to the bivariate normal form.

The pth order product moments of the joint distribution of $c(1)$ and $c(2)$ are sums of terms containing $$\Sigma' a_{gh} a_{ij} a_{kl} \dots \Sigma' b_{rs} b_{tu} b_{vw} \dots$$ or similar expressions in which arbitrary groups of suffixes within the Σ''s are tied, each Σ' involving products of p scores which may belong either to systems (1) or (2). Every such Σ' is in turn a linear combination of the corresponding Σ having the same suffixes and other Σ's in which additional tied suffixes are introduced. No Σ may contain a pair of free suffixes attached to one score, for it would then vanish by virtue of the fact that $\Sigma a_{ij} = 0$.

The even order product moments are first discussed. Let $p = 2m$. Consider a Σ in which the $2m$ scores are divided into m pairs each having one tied suffix, so that there are in all $3m$ independent suffixes, e.g. $$\Sigma a_{ij} a_{ik} a_{lr} a_{ls} a_{tu} a_{tv} \dots.$$

It may be written as $$(\Sigma a_{ij}^{(1)} a_{ik}^{(1)})^{\lambda} (\Sigma a_{ij}^{(1)} a_{ik}^{(2)})^{\mu} (\Sigma a_{ij}^{(2)} a_{ik}^{(2)})^{\nu},$$

where $\lambda + \mu + \nu = m$ and λ, μ, ν are the number of times the scores are paired in the combinations indicated.

As is always possible, suppose the numerically largest value of a_{ij} to be made equal to unity. We now impose the condition that $\Sigma a_{ij} a_{ik}$ is of order n^3 whether a_{ij} and a_{ik} belong to the same or different systems of scores. This is satisfied, when $\max a_{ij} = 1$, by τ and ρ, and also by r provided the sample is not an unusual one. With this condition, it is seen that Σ's of the above type are of order n^{3m}.

It is next observed that all other ways of tying suffixes give Σ's of lower order of magnitude. For the order of magnitude of the bracket is not reduced on replacing each a_{ij} by $+1$; consequently if further suffixes are tied the order of Σ is made less than n^{3m} since there are fewer than $3m$ summations from 1 to n. It follows that the dominant term in a Σ' is the corresponding Σ having the same array of suffixes.

Moreover, every non-vanishing Σ involving $3m$ independent suffixes can only be a permutation of the type illustrated, while those with more than $3m$ different suffixes must all vanish. This is made clear by considering how the $3m$ suffixes can be arrayed between the $2m$ scores. Begin by assigning $3m$ different suffixes at random among the $4m$ available places. At least m scores will receive their full complement of suffixes all which will be different. There cannot be more than m such completed scores, for if Σ is not to vanish, at least one suffix of each complete pair must be tied and this can only be done by repeating one suffix from every complete pair in each of the remaining places to be filled, of which there are only m. We are thus led to a permutation of the type of Σ discussed above. If there had been more than $3m$ different suffixes to begin with, there would not have remained sufficient empty places to prevent the existence of at least one score with a pair of free suffixes, and so all Σ's with more than $3m$ different suffixes must vanish.

Any $2m$th product moment is the sum of terms like $$\frac{(n-f)!}{n!} A \Sigma' a_{ij} a_{kl} \dots \Sigma' b_{rs} b_{tu} \dots,$$

where f is the number of independent suffixes in the Σ''s and A is a coefficient which is of

unit order as far as n is concerned. From the preceding argument, the maximum value of f is $3m$, in which case the term is of order $n^{-3m} \times n^{3m} \times n^{3m} = n^{3m}$. When $f \leqslant 3m-1$ the order of the term is not greater than $n^{-3m+1} \times n^{3m-1} \times n^{3m-1} = n^{3m-1}$ and such terms may therefore be neglected. Write

$$h_{11} = \Sigma a_{ij}^{(1)} a_{ik}^{(1)} \Sigma b_{lu}^{(1)} b_{lv}^{(1)}, \quad h_{12} = \Sigma a_{ij}^{(1)} a_{ik}^{(2)} \Sigma b_{lu}^{(1)} b_{lv}^{(2)}, \quad h_{22} = \Sigma a_{ij}^{(2)} a_{ik}^{(2)} \Sigma b_{lu}^{(2)} b_{lv}^{(2)}.$$

Then if terms of lower order of magnitude are neglected, the even product moment

$$\mu_{r,s} = \overline{c(1)^r c(2)^s}, \quad r+s = 2m$$

is given by the sum of terms like

$$n^{-3m} A_{\lambda,\mu,\nu} h_{11}^{\lambda} h_{12}^{\mu} h_{22}^{\nu}; \quad 2\lambda+\mu = r, \quad \mu+2\nu = s,$$

over all possible values of λ, μ, ν. The coefficient $A_{\lambda,\mu,\nu}$, which is the number of ways in which $h_{11}^{\lambda} h_{12}^{\mu} h_{22}^{\nu}$ can arise, is calculated as follows. Consider a Σ whose array of suffixes is such that it can be factorized as $(\Sigma a_{ij}^{(1)} a_{ik}^{(1)})^{\lambda} (\Sigma a_{ij}^{(1)} a_{ik}^{(2)})^{\mu} (\Sigma a_{ij}^{(2)} a_{ik}^{(2)})^{\nu}$. Its suffix pairs can be permuted within the sets of scores (1) and (2) in $r!\,s!$ ways, but of these $\lambda!(2!)^{\lambda}\,\mu!\,\nu!(2!)^{\nu}$ give essentially the same Σ. The suffixes within pairs attached to each score may also be rearranged in 2^{2m} ways without affecting the result, and so

$$A_{\lambda,\mu,\nu} = \frac{r!\,s!\,2^{2m}}{\lambda!\,\mu!\,\nu!\,2^{\lambda+\nu}} = \frac{r!\,s!\,2^{m+\mu}}{\lambda!\,\mu!\,\nu!}.$$

The calculation of the even order product moment $\mu_{r,s}$ for large n is in fact tantamount to selecting the coefficient of $t^r t^s / r!\,s!$ in

$$\frac{2^m}{n^{3m} m!} (h_{11} t_1^2 + 2h_{12} t_1 t_2 + h_{22} t_2^2)^m.$$

Finally, we dispose of the odd moments. In certain cases, such as for example the joint distribution of τ and ρ, they all vanish by symmetry. But even in the general case it can be shown that the odd moments are negligible to the order of magnitude $n^{-\frac{1}{2}}$.

A Σ containing $2n+1$ scores cannot have more than $3m+1$ different suffixes. For if there were $3m+2$, let them first be assigned to the $4m+2$ available places; at least $m+1$ scores will receive complete pairs of suffixes, and the remaining m empty places cannot be filled in any way which avoids one score having a free pair of suffixes. Hence as before the order of magnitude of any $(2m+1)$th moment is at most $n^{-(3m+1)} \times n^{3m+1} \times n^{3m+1} = n^{3m+1}$.

The $2m$th moments were shown to be of order n^{3m}, consequently if we define

$$\gamma(1) = n^{-\frac{1}{2}} c(1), \quad \gamma(2) = n^{-\frac{1}{2}} c(2),$$

the joint distribution of $\gamma(1)$ and $\gamma(2)$ has all its even moments of unit order, and by the result just proved all its odd moments are of order $n^{-\frac{1}{2}}$ and may therefore be neglected to that order. Reverting to $c(1)$, $c(2)$, it is seen that the moment-generating function of their joint distribution tends in the limit to the form

$$\exp \frac{2}{n^3} (h_{11} t_1^2 + 2h_{12} t_1 t_2 + h_{22} t_2^2).$$

Hence $c(1)$ and $c(2)$ tend to be normally distributed with variances $\frac{4}{n^3} h_{11}$, $\frac{4}{n^3} h_{22}$ and correlation

$$\frac{h_{12}}{\sqrt{(h_{11} h_{22})}} = \frac{\Sigma a_{ij}^{(1)} a_{ik}^{(2)}}{\sqrt{(\Sigma a_{ij}^{(1)} a_{ik}^{(1)} \Sigma a_{ij}^{(2)} a_{ik}^{(2)})}} \frac{\Sigma b_{lu}^{(1)} b_{lv}^{(2)}}{\sqrt{(\Sigma b_{lu}^{(1)} b_{lv}^{(1)} \Sigma b_{lu}^{(2)} b_{lv}^{(2)})}}.$$

The Γ's similarly tend to a bivariate normal distribution with the same correlation, but with variances,

$$\frac{4\,\Sigma a_{ij}^{(1)} a_{ik}^{(1)} \Sigma b_{lu}^{(1)} b_{lv}^{(1)}}{n^3 \Sigma a_{ij}^{(1)} a_{ij}^{(1)} \Sigma b_{lu}^{(1)} b_{lu}^{(1)}}, \quad \frac{4\,\Sigma a_{ij}^{(2)} a_{ik}^{(2)} \Sigma b_{lu}^{(2)} b_{lv}^{(2)}}{n^3 \Sigma a_{ij}^{(2)} a_{ij}^{(2)} \Sigma b_{lu}^{(2)} b_{lu}^{(2)}}.$$

Our proof rests on the assumption that $\Sigma a_{ij} a_{ik}$ and $\Sigma b_{lu} b_{lv}$ are of order n^3, where the individual a_{ij}'s and b_{lu}'s may belong to either score system. But if that is true, it follows that expressions like Σa_{ij}^2 must be of order n^2, for they cannot be made to exceed that order on replacing a_{ij} by $+1$, and their order cannot be less than n^2 since

$$\Sigma a_{ij}^2 - \frac{1}{n} \Sigma a_{ij} a_{ik} = \Sigma (a_{ij} - \bar{a}_i)^2 \geqslant 0,$$

where $\bar{a}_i = \frac{1}{n} \sum_{j=1}^{n} a_{ij}$. Consequently the variances of the Γ's decrease like n^{-1}. The correlation between the Γ's tends, however, to a value independent of n in the limit.

SUMMARY

The properties of a general class of correlation coefficients Γ, which includes the product-moment correlation coefficient r, Spearman's ρ and Kendall's τ, are discussed. A direct proof is given of the formula tentatively suggested by Kendall for the correlation between ρ and τ when the sample is permuted in all possible ways. The effect of a transformation of the sample values is also considered. It is shown that under certain general conditions, the joint distribution of two different Γ's, calculated on all possible permutations of the sample values, tends with increasing sample size to the bivariate normal form with variances inversely proportional to the sample size and correlation independent of it.

REFERENCES

HOTELLING, H. & PABST, M. R. (1936). Rank correlation and tests of significance involving no assumptions of normality. *Ann. Math. Statist.* **7**, 29.

PITMAN, E. J. G. (1937). Significance tests which may be applied to samples from any populations. II. The correlation coefficient test. *J.R. Statist. Soc. Suppl.* **4**, 225.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81.

KENDALL, M. G., KENDALL, S. F. H. & BABINGTON-SMITH, B. (1939). The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times. *Biometrika*, **30**, 251.

THE GROWTH, SURVIVAL, WANDERING AND VARIATION OF THE LONG-TAILED FIELD MOUSE, *APODEMUS SYLVATICUS*

By H. P. HACKER and H. S. PEARSON

I. GROWTH. By HELGA S. PEARSON

CONTENTS

	PAGE
1. Introduction	136
2. Trapping technique	137
3. Measurements	138
4. Monthly weight records	149
(*a*) General discussion	149
(*b*) Individual weight records. Growth graphs (Figs. 7–8)	151
(*c*) Combined weight records. Spot diagrams of month to month weight changes (Charts 1–3)	154
(*d*) Combined weight records. Monthly histograms (Figs. 10–12)	159
5. Summary	162

1. Introduction

Pioneer work in trapping field mice alive has been done in this country by Charles Elton, his colleagues and Dennis Chitty. Reference to similar work in other countries, especially in the United States, will be found in their papers and in the useful lists of recent literature in the *Annual Reports of the Bureau of Animal Population, Oxford University.* We record with gratitude the benefit we have derived from their labours and devices.

Our own interest lay primarily in comparing, by simple statistical methods, populations from species of mammals distributed more or less continuously over great distances. These comparisons were to be based on skeletal measurements and proportions. We were limited to our own country and to a ubiquitous mammal easily caught in large numbers; thus field mice seemed to be the only choice, although for skeletal measurements these small rodents present great difficulties. Our Oxford advisers recommended the long-tailed field or wood mouse, *Apodemus sylvaticus sylvaticus*, as more or less ubiquitous and easy to catch. We decided to follow in their footsteps, and start by trapping these mice alive, marking them, setting them free and trapping again, in the hope of getting some first-hand experience of how far they wander, and whether they are confined to particular types of country or can be regarded as a single population continuous throughout England.

Even if a whole species can be regarded as a continuous population it is clear that there can be no such thing as 'random mating' within it, if the individuals rarely wander more than a few hundred yards. There must be regional inbreeding, though there may be no boundaries between the regions. Can any regional differences be detected between field mice if measurable characters are compared by statistical methods?

The trapping of *Apodemus* alive is a fascinating pursuit. Many more questions arise from it than can be answered perhaps in a lifetime, and it has proved hard to hold firmly to the pursuit of our original problem. In trapping over a number of seasons it has been easy to collect a vast quantity of facts, and these, when arranged and digested, cry out for the

collection of a vast quantity more. It has seemed advisable, however, to try to publish some of them at this stage, in the hope that they may prove of interest or use to other workers, if we ourselves are not able to follow them up much further.

The data may be arranged, though with much overlapping, under the four headings: Growth, Variation, Wandering, Survival. We hope to publish some selected facts appropriate to these headings in a short series of papers of which this is the first.

2. TRAPPING TECHNIQUE

That many of our present records are not suitable for statistical treatment is due to our ignorance when first trapping of the best method of laying out our traps. Only now, after five seasons' trapping following on the work of Elton and of Chitty, are we beginning to understand the conditions under which large numbers of *Apodemus* can best be caught.

There are four main limiting factors:

(i) *Man hours.* Elton *et al.* (1931, p. 661) stressed the number of 'man hours' needed for setting, visiting, and resetting large numbers of traps, and for dealing with the mice that are found in them. The largest number of traps we ourselves have ever set out at a time was 96, and 48 mice was our largest night's catch, of which 4 were bank voles and 44 *Apodemus*. Our method was to pick up all the occupied traps, replace them with spare traps, and take them home. There we identified or marked the mice, weighed and measured them, placed them in cages, and wrote up our day's records. Then we rebaited the traps and renewed the food in the attached nest boxes (see p. 145), so that they were ready for taking out to replace others next day. Our biggest catches meant a long, intensely hard day's work for three practised people; had there been more of us we could have put out more traps covering a wider area and caught more mice, and so have had more adequate data for statistical treatment.

(ii) *Weather.* Setting out a large number of traps does not necessarily mean catching a large number of mice. On some nights we have had all our traps out in likely places without a single catch. Weather conditions undoubtedly play an all important part in keeping mice from wandering freely, and this in itself needs a carefully planned investigation. We have learnt from experience that a night of blustering south-west wind is a good trapping night and that, if snow is on the ground, it is of little use leaving out the traps when once the mice living in their immediate neighbourhood have been caught; but we are not sure of the relative effect on the numbers caught of wind, moon, temperature, rain or hoar frost.

(iii) *Arrangement of traps.* Only a certain number of mice will be within reach of any one trap under any particular set of weather conditions, and a great number of traps in a dense 'scatter' will catch no more mice under those conditions than few traps more widely spread. A line of traps will drain a wider area than an equal number arranged in a grid, while a hollow square may catch all the mice within it just as easily as a grid covering the same area. With experience some idea may be gained of a good arrangement, but this must still depend to a large extent on a further limiting factor:

(iv) *Habitat.* Not all places are equally frequented by *Apodemus*, yet this cannot easily be attributed to obvious ecological differences. We have caught mice in large numbers in very varied habitats, whereas a comparatively uniform plant community may have good and bad patches. We are inclined to suspect that any kind of overhead cover, such as dense bracken, heather or thorny scrub, is of as great importance as food or soil factors, but much more work is needed to discover where the mice live as apart from where they wander, and

whether or no they live in large communities in the winter and separate before the breeding season. Light might be thrown on food preferences by analyses of stomach contents and by experimental feeding.

The little information we have on all these matters, together with further notes on our methods of trapping and marking, we hope to publish in our papers on survival and wandering.

3. Measurements

For our future comparison of populations from different localities we needed the skeletons of fully grown mice. In order to find out at what time of year a mouse population contains the largest proportion of such mice, we studied the growth changes of our living population, weighing and measuring each mouse, and recording the numbers of mice of each size in each month; we were also able to keep individual growth records of marked mice. The data thus collected are analysed in the last three sections of this paper.

We had often to handle a large number of live mice in a very little time, so we limited ourselves in their case to two comparatively easy measurements: (i) the weight, (ii) the length of the right hind foot. For any mouse found dead in a trap, or that died in captivity, we made further records of: (iii) the total length of head and trunk, (iv) the length of tail, (v) the weight of stomach, (vi) the weight of the reproductive system in the male, or of visible embryos in any pregnant female.

We have made these same measurements on all mice trapped for skeletons in other localities, and have so gained a first rough index to regional differences. So far we have carried out this regional trapping near Westerham in Kent, Hampden in Bucks, Swanage in Dorset, and St Mawes in Cornwall. We trapped in each of the years 1937, 1938 and 1939, at the very end of March or beginning of April, a time when winter males are past their period of maximum weight increase (p. 155), and when young mice are in most years soon likely to appear in the traps. These dead mice together form the adult population from which the data of Table 1 (p. 139) were calculated. In Fig. 1 their measurements are added to those on dead mice caught at Holwood Park and Downe in Kent, and at Cobbler's Hill in Buckinghamshire, at other times of the year. In a future paper we hope to give an analysis of the local variation shown by this material, together with such further information on growth as can be gained from correlating the external and skeletal measurements of the series of dead mice of all ages from Holwood.

To facilitate statistical analysis the records of each mouse, dead or alive, were entered straightway on to separate index cards, the reference number on each card being that of one individual mouse.

Live mice. It was easy to weigh the live mice. We let them drop from the trap into a large net bag, and from there transferred them by hand to a small, narrow bag in which they could be tied firmly and placed on the balance. The feet were not quite so easy to measure accurately, but we coaxed the mice by gentle pressure from the small bag into a cupped hand, one of us holding them there with the right-hind leg projecting between forefinger and thumb, while another (always the present writer) stretched the hind foot at right angles over the holder's forefinger and measured it from heel to fleshy tip of central digit with a pair of screw callipers with Vernier scale. In this technique we attained a degree of accuracy represented by a standard error for a single measurement of 0·18 mm.*

* This is the S.D. for repeat measurements on the same mouse by one observer, based on 10 or more observations on each of 50 adult male mice, the average foot-length being 23·0 mm.

We sexed each mouse while holding it, and noted whether a female's vulva was perforate. We also placed it in a glass inspection jar—a barrel-shaped lamp glass with one opening sealed and the other wide enough to admit a hand—and noted the amount of yellow streak on the hairs of the chest, a very variable feature.

Post-mortem errors. That favourite measurement of systematists, the combined length of head and trunk, is subject to very great observational error. Indeed, a wide range of error must be allowed for in all external measurements on dead mice, taken as they usually are at varying intervals after death.* Loss of body weight due to drying or decomposition takes place slowly, and we found no appreciable change in 24 hr., but the feet tend to dry and shrink before this. Length of foot, and length of head and trunk, will both be shorter if measured during the period of *rigor mortis* than either before or after; they vary with the stage of rigor and the amount of massage given before measuring, though in extreme rigor it is doubtful if the massage has much effect. As far as dead mice are concerned we hope to obtain more accurate information about growth from skeletal measurements, since these are of a more stable nature and less liable to observational error.

Table 1. *Above: The variation in five measurements on a number* (N.) *of dead adult or nearly adult* Apodemus *caught in four localities between* 28 *March and* 13 *April* 1937, 1938 *and* 1939. *Mean* (M.), *standard deviation* (S.D.), *and coefficient of variation* (C.V.). *Below: The correlation between these measurements; correlation coefficient* (C.C.).

Measurement	Male				Female			
	N.	M.	S.D.	C.V.	N.	M.	S.D.	C.V.
Weight (g.)	161	23·4	2·38	10·2	115	18·8	2·20	11·7
Head and trunk length (mm.)	162	88·6	4·84	5·5	121	84·1	4·89	5·8
Tail length (mm.)	142	90·3	5·27	5·8	107	86·3	4·96	5·7
Right hind foot length (mm.)	162	22·9	0·82	3·6	121	22·3	0·69	3·1
Weight of reproductive organs (g.)	149	2·0	0·41	20·5				

Correlation	Male		Female	
	N.	C.C.	N.	C.C.
Head and trunk length with weight	161	0·59	115	0·53
Head and trunk length with tail length	142	0·51	107	0·46
Head and trunk length with hind foot length	162	0·56	121	0·47
Tail length with hind foot length	142	0·57	107	0·38
Weight with weight of reproductive organs	149	0·66		

Weight as a measure of growth. The easiest measurement of a live mouse, its weight, is unfortunately not a very satisfactory measure of growth. Apart from seasonal changes, it alters considerably from one day, or even hour, to another. This must be partly due to the condition of the stomach and bladder, but we suspect that a rapid fall in actual body weight

* Sumner (1927) has shown how great this error may be, by comparing measurements made on the American deer-mouse *Peromyscus*, (*a*) by different observers, (*b*) by the same observer at different times.

can also occur after a period of prolonged activity. Frequent trapping, and trapping during the breeding season, may also interfere with weight increase.

Some idea of the extent to which the weight of a mouse is correlated with its length, and therefore indirectly with its growth, is given by the scatter diagram of Fig. 1. We were not able to measure the length of our live mice, but this diagram is based on the measurements of the dead male mice referred to on p. 138. It must be examined with reserve, however, as the number of young is few, and the mice were caught at varying times

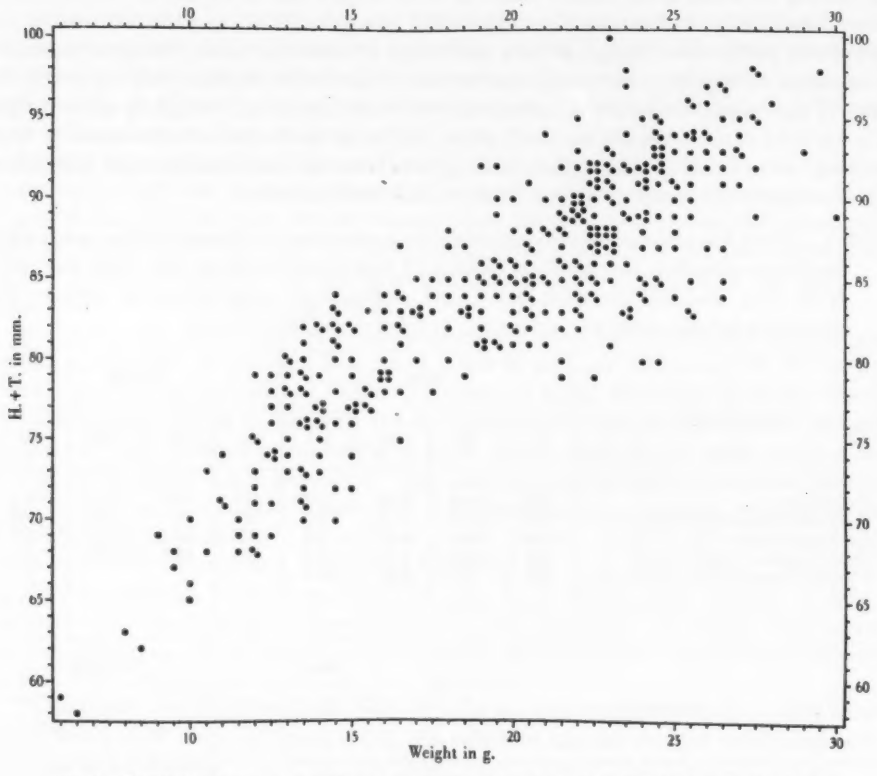


Fig. 1. Correlation between weight and length of head and trunk in 367 dead male *Apodemus* caught in various localities at various times of year.

of year in several widely separated localities, the young in different localities from most of the adults; and, as we hope to show in this and a later paper, weight varies with season and size with locality independently of the age of a mouse. In the mixed batch of *adult* or *nearly adult* mice from Westerham, Hampden, Swanage and St Mawes (included in the upper part of the diagram), we have found the coefficient of correlation between these two measurements to be $r = 0{\cdot}59$ for 161 males and $r = 0{\cdot}53$ for 115 females (Table 1). To calculate comparable coefficients for growing mice would call for far larger numbers in the lower weight groups, and for this reason, and because of the mixed nature of the material,

we have thought it inadvisable to fit a curve to the diagram.* Any such curve would tend to flatten as growth ceases, increase of length with increase in weight becoming solely due to the size variability of the fully grown population. For apart from all the sources of variation mentioned, weight and length must vary not only with age but also with individual character. We have been unable to separate these two sources in growing mice, and without more frequent individual records (and these would probably interfere with growth) it does not seem possible to be certain that any one live mouse is fully grown. Records taken throughout the year suggest that, in the single locality of Holwood Park, 20·0 g. in a male and 17·0 g. in a female can be taken as arbitrary but useful minima below which most of the mice may be expected to be still growing, except in midwinter when heavier adults often fall to lower levels than this (p. 154). It will be seen from Fig. 1 that males of over 20·0 g. range from under 80 to 100 mm. in length, while those of less than 20·0 g. can range as high as 90 mm.

A measure of the variability of weight in *adult* mice from scattered localities is given by the coefficients of variation (100 × standard deviation/mean) which we obtained from the mixed batch mentioned above. These coefficients are given in Table 1, where they are seen to be nearly double the coefficients for length of head and trunk, although the variability in this case includes much greater observational errors. Ruger (1933) found coefficients for weight as high as 15·93 and 14·77 in his human material (corrected to the standard ages of 40·5 years for men and 32·5 years for women), while his coefficients for stature, sitting height, and span are less than ours for head and trunk length, and for tail length.

Summary of factors influencing weight.

(1) Age.
(2) Inherent† size of mouse.
(3) Locality.
(4) Season: (i) fat deposit; (ii) condition of reproductive organs and, in females, presence of embryos.
(5) Prolonged activity.
(6) Trapping interference.
(7) Content of stomach and intestines.
(8) Content of bladder.
(9) Disease (see footnote to p. 158).

Tail length. The length of the tail from anus to tip, excluding the terminal pencil of hairs, is another conventional record of systematists. It is more easily taken on a dead animal than is the length of the head and trunk, and is less subject to observational errors, but as the tail is liable to shortening through injury, our number of records is not so great. The average length of tail in adult mice is very similar to the average length of head and trunk, and this is true for mice of all ages found in the traps, indicating that the tail continues to lengthen as long as the body is lengthening. On the other hand, each varies

* A clear analysis of the difficulties of fitting curves to comparable but much more numerous human measurements has been given by Ruger & Stoessiger (1927) and Ruger (1933). From measurements of fifteen characters taken in 1884 by Francis Galton on over 7000 men, women and children, they computed means, standard deviations, correlation coefficients, correlation ratios and regression equations, and constructed graphs of the regression lines. These papers, and the analytical methods of the biometric school in general, are well worth the consideration of those biologists who attempt to base systematic classification, and even evolutionary theory on so-called *allometric growth curves*.

† *Inherent.* Def. 2 in *Shorter O.E.D.*: 'Existing in something as a permanent attribute or quality'.

to a considerable extent independently of the other (for adult mice, $r = 0{\cdot}51$ in males and $r = 0{\cdot}46$ in females), so that neither can be taken as an index to the other in any particular mouse. Fig. 2 is a correlation diagram for the two characters derived from the same series of dead male mice that formed the basis of Fig. 1. The same qualifications, due to heterogeneity of the data, apply as were referred to on p. 140.

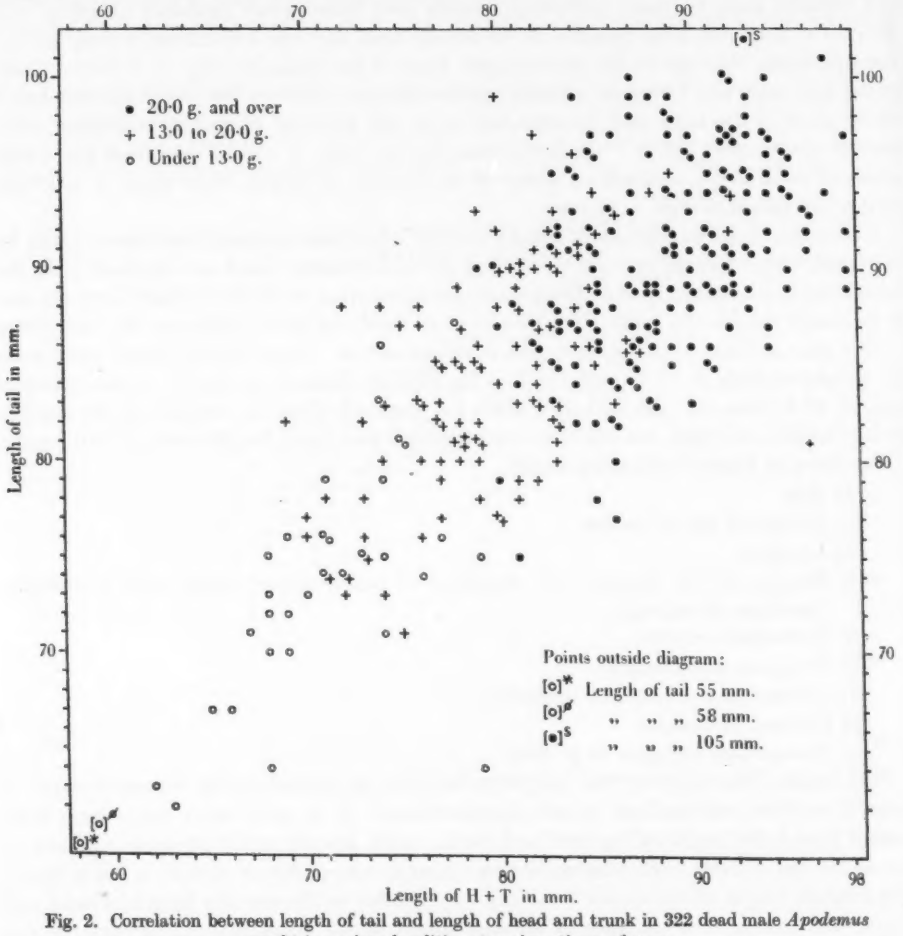


Fig. 2. Correlation between length of tail and length of head and trunk in 322 dead male *Apodemus* caught in various localities at various times of year.

An exceptionally short tail in any weight group may indicate an early loss of tip not perceptible or not noticed on measuring. If *Apodemus* is caught by the end of the tail, the skin easily slips off, if the grip on the bone is not tight, often allowing the mouse to escape; the protruding bone soon dries up and breaks off. If the loss is of any length—and more than half the tail may be lost in this way—it is easily perceptible, as the tail ends bluntly without the terminal pencil of hairs, and sometimes with a short length of exposed bone; in such cases no measurement was made.

Length of hind foot. From our measurements on live *Apodemus* we have found that the hind foot reaches maximum length long before the mouse stops increasing in weight, so that this length, within the error of measurement (see p. 138), early becomes a fixed

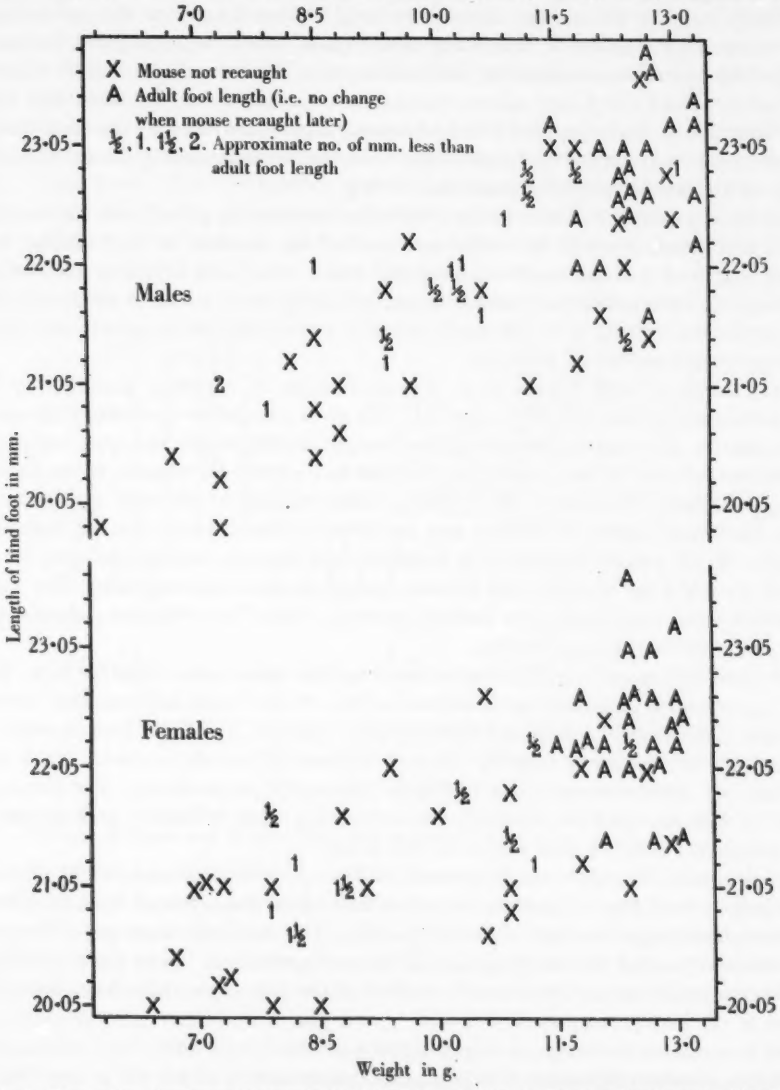


Fig. 3. Young live *Apodemus*, Holwood Park, 1937–8 and 1938–9. To show that length of hind foot rarely increases after a weight of 12·0 g. is reached in males, or 11·5 g. in females.

characteristic of the mouse, and is especially useful in establishing local differences. Fig. 3 shows that the hind foot of the Holwood Park mice ceased to grow after the mice had reached a weight of 12·0–13·0 g., about half the maximum weight of the males. Where

the mouse was caught only once the measurement (marked by a cross) throws no direct light on the final value of the foot length, but it combines with the rest of the data to give a scatter diagram indicating that foot length was still increasing in growing mice at the lower body weights. Where the mouse was caught more than once the measurement is either marked by a letter *A*, indicating adult value, when no subsequent increase was found, or by a number, representing the number of millimetres short of adult value. It is clear that in almost every case where repeated measurements were obtained four or more weeks later (see p. 149), the hind foot had ceased to grow in length by the time the mouse weighed 13·0 g., and very often before that. Thus this measurement gives no indication of the age of any mouse weighing more than 13·0 g.

In an April population (Table 1) the correlation between length of hind foot and length of head and trunk ($r = 0{\cdot}56$ for males and $r = 0{\cdot}47$ for females) is very similar to that between length of tail and length of head and trunk; that is to say, it is not sufficiently great to give a clear indication whether an exceptionally small mouse in such a supposedly adult population is likely to be inherently small or merely delayed in growth and therefore to be rejected for statistical purposes.

Table 1 *compared with human data.* The coefficients of variation obtained by Ruger (1933) have already been referred to (p. 141). He gives correlation coefficients for span and weight, stature and weight, sitting height and weight, sitting height and span, ranging from $r = 0{\cdot}554$ to $r = 0{\cdot}598$ for men, and from $r = 0{\cdot}434$ to $r = 0{\cdot}595$ for women, those for women being quoted from Elderton & Moul (1928); these coefficients are very similar to those for the *Apodemus* material of Table 1 and, as there, are less in every case for females than for males. Much greater correlation is found in man between stature and span ($r = 0{\cdot}818$ for men, $r = 0{\cdot}824$ for women) and between many skeletal measurements. The material from which these coefficients were derived, however, was of too different a character from ours to warrant detailed comparison.

A reference to Sumner's (1926) data on American deer-mice is also called for here. Besides other measurements and colorimetric estimates, he gave the means and standard deviations for weight, body length (= head and trunk length), and foot length, in three varieties of the *Peromyscus polionotus* group from Florida and Alabama. These are evidently much smaller mice than our *Apodemus sylvaticus*, and quite differently proportioned. For geographical reasons his data are again not directly comparable with those in Table 1, and comparison is better postponed until we deal with local variation.

Stomach weight. The part which stomach content plays in the total weight of a mouse can be judged from Fig. 4, the data for which came from the series of dead mice trapped with break-back traps for their skeletons (p. 138). The stomachs were cut at the pylorus and at the lower end of the oesophagus, and the spleen removed. As no material difference could be detected between the stomach weights of the two sexes these have been bulked together in the histogram.

It will be seen that the range of weight in 268 mice was from 0·3 to 4·0 g., with a mean of 1·2 g. and a standard deviation of 0·79 g., but that stomachs of 0·3–0·8 g. were the commonest. Weights above 3·0 g. were rare, and the one 4·0 g. stomach, that of a female, was so enormously distended that it appeared to fill the whole abdominal cavity. When the stomachs of two male and two female mice were scraped clean of contents, one of them weighed 0·3 g. and the others approximately 0·25 g., so that the ten stomachs in the 0·3 column in Fig. 4 may be regarded as empty. For younger mice the upper limit of stomach

weight will of course be less, varying with the size of the mouse, while the lower limit may be as little as 0·1 g., as we have found in the unscraped stomach of a 7·0 g. youngster.

The prevalence in the snap traps of mice with empty or nearly empty stomachs indicates that usually only hungry mice seized the bait. Mice living near the site probably found the traps early in their nightly wandering in search of food; they rarely had time to swallow the bait, though they may have eaten some of the oat grains scattered in front of it as an extra attraction.

But the mice with whose growth this paper is mainly concerned were caught alive in large box traps,* to the end of which was fixed a ½ lb. cocoa tin with bedding, two peanuts, and a large pinch of oats. Together with the bait (another peanut), this was more food than they usually ate, so that when weighed these live mice are less likely to have had an empty stomach than those caught in snap traps. On the other hand, with excess of food available, the stomach content may have depended on the time of day at which they were taken from the traps to be weighed, and this varied from 10 a.m. to 7 p.m. according to

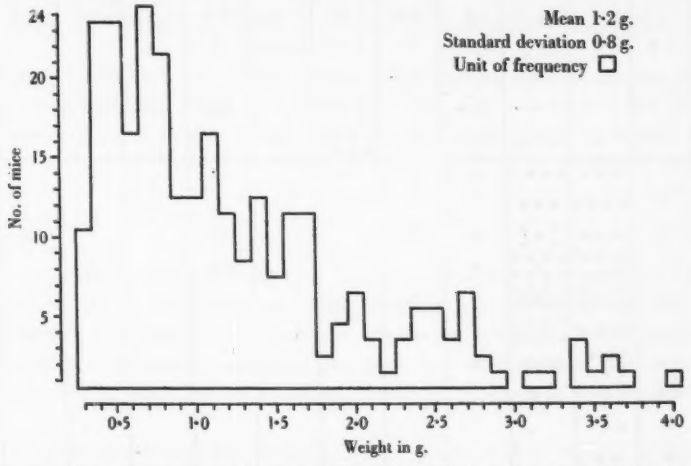


Fig. 4. Frequency distribution of stomach weight in 268 male and female *Apodemus* caught in four localities between 28 March and 13 April 1937, 1938 and 1939. 0·3 g. is the weight of an empty stomach.

the size of the previous night's catch. Elton *et al.* (1931, p. 716) recorded some observations on the times of feeding activity in *Apodemus*, and on the quantities eaten, but more data about this are needed. All we can say here is that a variation of as much as 4·0 g. in the weight of a full-grown mouse *may* be due to stomach content, while any variation of up to 1·5 g. must so commonly be due to this factor that it should be discounted as an indication of change in body weight *where only a single mouse is concerned*.

Reproductive organs. The growth of the reproductive organs in the male, and the presence of embryos in the uterus of the female, are among the factors influencing body weight. We have dissected out and weighed together the testes, vesiculae seminales, Cowper's glands and penis of all dead male *Apodemus*. We have also weighed together in their uterine casing any embryos large enough to make visible swellings in a dead female. Some of these mice

* Selfridge traps (Elton *et al.*, 1931, p. 14). We are indebted to Messrs Selfridge for supplying us with these traps at a reduced rate.

were found dead in the box traps set in Holwood Park, and some were caught in snap traps elsewhere.

(*a*) *Male.* The weight changes of the reproductive organs of a male mouse appear to be correlated with the changes in its body weight. We shall presently show that in midwinter only very small males gain appreciably in body weight, while larger mice are stationary, and autumn adults lose weight (pp. 158–161 and Chart 1, C1, D1, D2). Fig. 5 shows the weight of the reproductive organs of over 200 male mice of varying weights in different

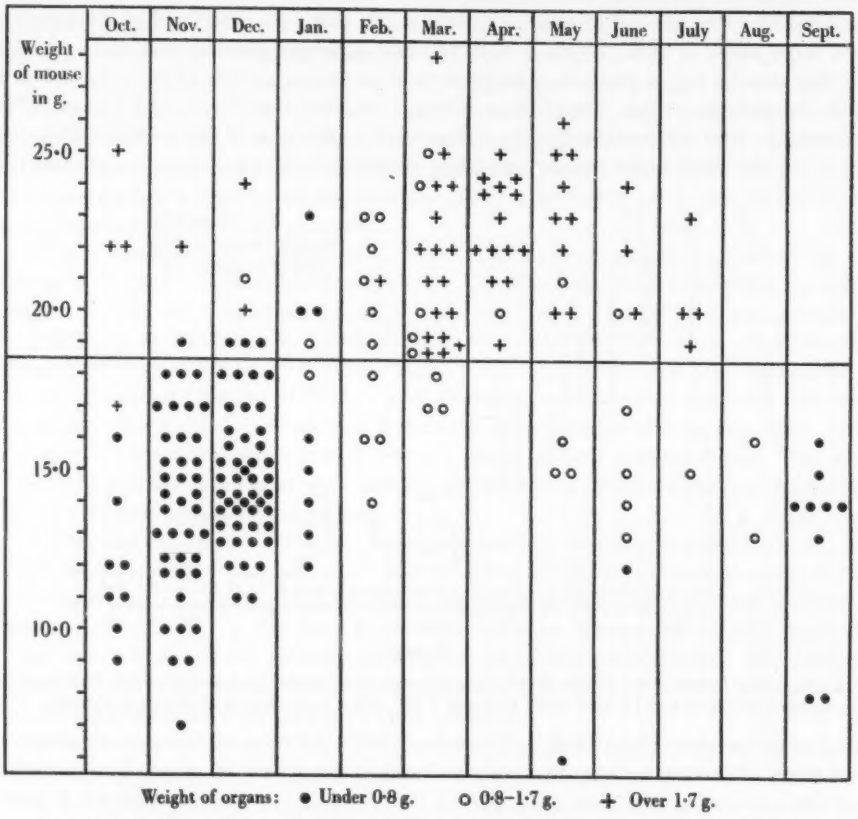


Fig. 5. Weight of reproductive organs in male *Apodemus* of different weights at different times of year. The mice were caught in several years and several localities.

months of the year. Out of 112 of these mice caught in the months September to January inclusive (the majority in November and December), only fifteen weighed more than 18·5 g., the mean weight being 14·5 g. as in the November and December histograms of live mice on p. 159. Only six of these fifteen mice of adult or nearly adult size had well-developed reproductive organs weighing more than 1·7 g., and only two had organs of intermediate weight; while the remaining seven, and all but two of the mice of under 18·5 g., had organs weighing less than 0·8 g., by far the most frequent weights being from 0·1 to 0·3 g. In

several mice weighing from 17·0 to 19·0 g. the testes had evidently been larger, but were withdrawn into the abdomen and appeared like empty, shrunken sacs; the vesiculae seminales and Cowper's glands were also shrunken. These mice may have already reached maturity, but had undergone loss of weight and of reproductive activity at the onset of winter, and were in a condition similar to that described by Brambell & Rowlands (1936) in bank voles.*

After January, when our live records show that the majority of male mice gain in body weight (pp. 154, 155 and Chart 1, C3), it appears that their reproductive organs also grow rapidly in most years, coming into the intermediate group of 0·8–1·7 g. The eleven February mice shown in Fig. 5 were all in this stage except one, in which the organs already weighed 2·0 g. Among other mice in this stage were:

(i) A single December mouse of 20·5 g., whose testes were still unshrunken though small.

(ii) Two advanced January mice.

(iii) Several March mice, presumably slow in their development.

(iv) A group of young summer mice, all under 17·0 g.

(v) Three large mice caught in April, May, and June, of whose significance we are in doubt.

All but one of the 55 mice with reproductive organs weighing over 1·7 g. were themselves over 18·5 g., and therefore of adult or nearly adult size; the majority of the March mice had reached this stage, the weight of their reproductive organs ranging as high as that of adult mice in later months.† Since the range of body weight of adult spring and summer mice is from just over 18·0 to just over 28·0 g., and the range of their reproductive organs is from 1·8 to 2·8 g., if there is any correlation between the two weights it appears that in spring and summer about 10 % of the body weight may be attributed to the reproductive organs. This is very different from what may be expected in midwinter, when these organs contribute only 0·1–0·3 %. We have very little data for any one month in the summer, however, and cannot trace the month to month connexion between the two weights. There is some slight evidence from our records of live mice that a temporary loss in body weight occurs after breeding (p. 155), and it would be of interest to know if this is reflected in the reproductive organs.

For April, a better picture of the range in body weight and in weight of reproductive organs and of the correlation between them ($r = 0{\cdot}66$), is given by the scatter diagram in Fig. 6. This represents 149 male mice of which the majority were probably adult; they are among those described on p. 138 as caught in snap traps between 28 March and 12 April in four different localities in 1937, 1938 and 1939.‡

(*b*) *Female.* Of 121 females caught at the same time as the April males just referred to, thirteen had embryos visible in the uteri, indicating that the breeding season had started in all three years, though there were actually some local differences in this respect. These

* We understand that these authors have material for a similar paper on *Apodemus*; much of our own work must be complementary to theirs, and we very much regret that they have not yet been able to publish it.

† Nineteen males were caught in February and March 1942, too late to be included in Fig. 5. In this year both January and February were exceptionally cold months; the mice caught in February corresponded in body weight, and in the development of their reproductive organs, with those caught in January in previous years, while those caught in March corresponded with February mice of previous years.

‡ Since the four groups of mice from these four localities differ among themselves and from the Holwood mice in size (including that early maturing element hind foot length), and therefore in body weight, the rough estimate of 20 g., given on p. 141 as the weight below which most Holwood males may be expected to be still growing, cannot be applied to these data.

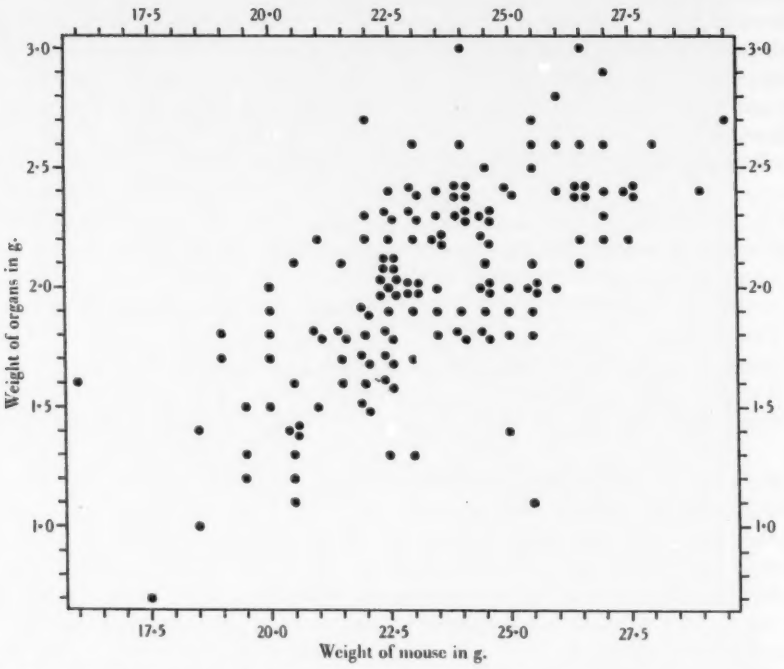


Fig. 6. Correlation between weight of mouse and weight of its reproductive organs in 149 male *Apodemus* caught in four localities between 28 March and 13 April 1937, 1938 and 1939.

Table 2. *Female* Apodemus *caught between* 28 *March and* 13 *April in* 1937, 1938 *and* 1939, *in four different localities. Weight distribution of those with and without embryos.*

Weight of mouse to nearest g.	No. without visible embryos	With visible embryos			Total no. of cases
		No. of cases	No. of embryos	Weight in g. of embryos and uterus	
15·0	5	—	—	—	5
16·0	8	—	—	—	8
17·0	27	—	—	—	27
18·0	16	1	5	0·6	17
19·0	16	1	4	1·1	17
20·0	12	1	4	*S*	13
21·0	14	4	?, 4, ?, 4	0·3, 0·3, *S*, *S*	18
22·0	6	—	—	—	6
23·0	2	4	3, 4, ?, 6	4·5, 1·8, *S*, *S*	6
24·0	2	1	4	1·5	3
⋮					
28·0	—	1	4	6·6	1
	108	13			121

females ranged in weight from 14·5 to 24·5 g., with an outlier at 28·3 g. They are shown in Table 2, which gives particulars of those with visible embryos; where the letter *S* is recorded in the last column but one the weight for embryos and uterus was negligibly small compared with the body weight.

It is clear from these data that even at the beginning of the breeding season, heaviness in a female is not necessarily due to the weight of embryos. In our live trapping in 1939 we found that there was a period of universal weight increase for females between March and April, when mice of all weights showed a wide range of increase (p. 158, and Chart 2, C5, C5X, D5, D5X), although the weather was very cold and no young appeared in the traps until May: there seemed no connexion between pregnancy and previous weight. Prof. Brambell's material may throw more light on this point. In non-pregnant females the reproductive system contributes but an insignificant amount to the weight of the mouse.

4. MONTHLY WEIGHT RECORDS

(*a*) *General discussion*

Our best records of weight changes in live mice are from the season 1938–9. We trapped each month in two neighbouring parts of the Holwood Park woods,* a week in one part followed by a week in the other, from the beginning of the third week in November until the summer breeding season.

We have given the weeks of the year arbitrary numbers, so that one year may be the more easily compared with another. Week 1 always starts on the first Monday in October. This has the disadvantage that there may be as much as six days' difference between a week in one year and the week with corresponding number in another year.

Month	West of footpath		East of footpath	
	No. of week	Dates	No. of week	Dates
November	7	14th to 17th	8	21st to 25th
December	11	12th to 15th	12	19th to 23rd†
January	15	9th to 12th	16	16th to 19th
February	19	6th to 9th	20	13th to 16th
March	23	6th to 11th	24	13th to 18th
April	29	17th to 22nd	30	24th to 29th
May	33	15th to 18th	—	—
June	—	—	38	19th to 22nd
July	43	26th to 27th	—	—

It will be seen from this list that trapping was repeated every four weeks until after the March catch, but that six weeks elapsed between the beginning of the March catch and the April catch, five weeks from May to June, and five weeks from June to July. In the first four months we usually trapped for four nights only in each area, the first four nights of each trapping week; in March and April for six nights in each trapping week, in May and June for four nights, in July for only two nights. In each of the last three months we trapped in only one part of the woods in order to interfere less with breeding.

Chitty (1937, p. 43) has shown that mice caught one night and set free next day are

* We are greatly indebted to the late Lord Stanley and to Lady Stanley for permission to trap on the Holwood estate at Keston in Kent, the site of a great Iron Age camp and once the home of William Pitt, who here enclosed many acres of common land, tended the woodlands, and planted trees on the bulwarks.

† All traps were taken in on the night of 21 December owing to a fall of snow.

likely to reappear in the same or a neighbouring trap night after night. This keeps other mice out of the traps, or, if surplus traps are set, greatly adds to each day's labour. We therefore placed all the mice caught in cages, a separate cage for each trapping site, and released them in their home area only after the traps had been taken up at the end of the week's trapping. An exception was made in the case of pregnant or lactating females, which were set free each day; even these generally came back to the trap next night, and we are afraid that some interference with breeding must have taken place. At times young were born in the traps, or before the mother could be set free, and she then generally killed and ate them, or else left them to die. Once when we set a mother free with her young where caught, she did not take them back to the nest, but they were found abandoned and dead on the same spot next morning. At other times a released female would show a great drop in weight when recaught next day, suggesting that her family had been born in the interval (p. 153 and Fig. 8, group VI, no. 600); as she may have been away from the nest for more than 12 hrs. it is all too likely that the nest young died meanwhile.*

Even out of the breeding season, during a period of active growth, there seems some evidence that frequent catching interferes with weight increase. Mice caught in two consecutive weeks, as happened frequently to those living on the adjoining borders of our two areas, often weighed less at the second catching. We have further evidence of this from other years when one area was trapped again after a short interval. It is conceivable that the commencement of breeding in a whole population might be delayed by too frequent trapping and too long periods of captivity.

Chitty (1937, p. 46) has given a list of eight disturbing effects of trapping, and our method of keeping the mice together in cages for several days is likely to enhance and add to these. On the other hand, it probably enabled us to catch a greater number of mice and thus to have better data for statistical treatment; for we have evidence that the local inhabitants can be removed from an area in a night or two, and that outsiders then wander into it, if suitable weather conditions prevail. This evidence we hope to give in a future paper on wandering.

We caught 343 different mice in the season 1938–9, of which 134 were never recaught, including nine which died in the traps or cages. Twenty-three mice were under 10·0 g. in November and December, but only nine of these were ever caught again, and only four reappeared each month throughout the season. These four mice gave us our most complete individual weight records, which are graphically represented in Figs. 7 and 8, groups I–VI, together with those of a number of other mice of greater weight when first caught. In these graphs as many mice have been selected as could be represented clearly in a few text-figures. Mice of less than 7·0 g. rarely appear in traps, and it is probable that they do not wander out from the nest until about this size unless the nest is destroyed or their mother fails to return to it; the smallest mice we have ever caught in traps were two males of 5·5 and 5·8 g.

Besides these fairly complete individual records we had many lasting over shorter periods. The spot diagrams of month to month weight changes (Charts 1–3) and the monthly weight frequency histograms of Figs. 10 and 11, are an attempt to utilize these data by massing them, which should to some extent smooth out hour to hour variability due to such causes as change in stomach content.

* Burt (1940, p. 15) has, however, 'one record of newly born *Peromyscus* which lived sixty hours without parental care of any kind'.

(*b*) *Individual weight records. Growth graphs*

Males. Fig. 7, group I, shows the rapid rise in weight of four young males between the November and December catches. The rise was greatest in the smallest mouse, but not sufficient for it to overtake the others. Between December and January the rise was very slight; in the smallest mouse, no. 468, it was still slightly greater than in nos. 423 and 528,

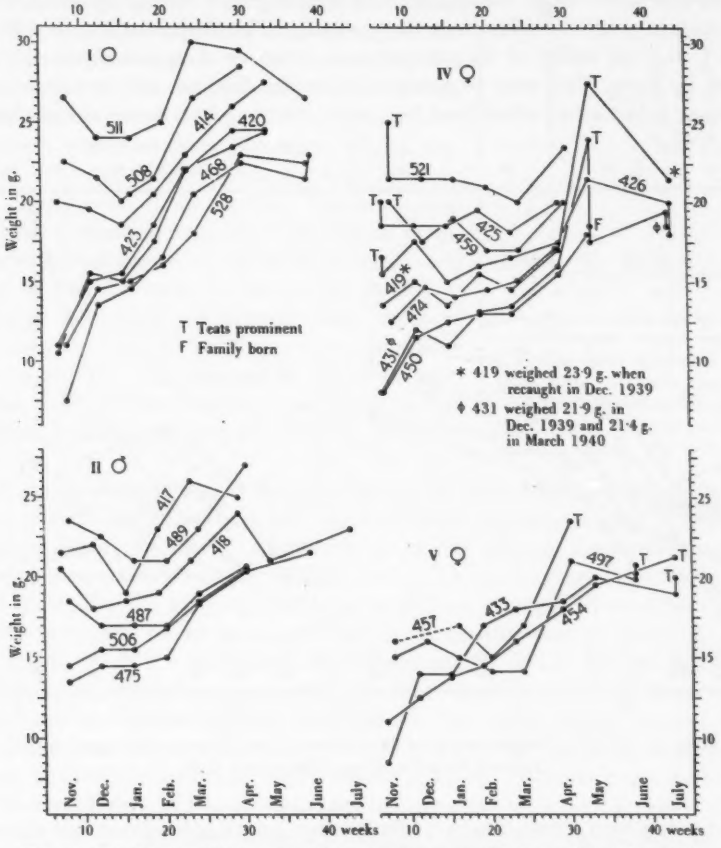


Fig. 7. Individual weight records of live *Apodemus* caught at monthly intervals in Holwood Park, November 1938 to July 1939.

whereas no. 420 actually lost a little. At the top of group I, nos. 508 and 511, and possibly also no. 414, were already of adult size in November; they all showed loss of weight between November and December, and the first two continued to drop until January, while no. 511 remained stationary. All the mice began to put on weight after the January catch, the older mice still remaining ahead of the younger mice when the breeding season was reached. Taking them all together the period of most rapid growth was between February and March; the increase tended to slacken after March or April, and was followed in three of them by actual loss in weight.

Fig. 7, group II, shows three more mice of adult size losing weight between November and January: nos. 489, 417 and 418. No. 487, of only 18·5 g. in November, also lost weight, and may also have reached maturity in the autumn breeding season, but if so it was an exceptionally small mouse. Nos. 506 and 475, of 14·5 and 13·5 g. in November, were evidently older than the young mice in group I and showed a much smaller increase than these by December; they did not change between December and January, and again showed smaller increases in the spring, both reaching only 20·5 g. by April, whereas the younger mice in group I reached from 22·5 to 24·5 g. The older mice in group II, like those in group I, kept far ahead of the younger ones, with the exception of no. 487 which was overtaken by them. This may be another indication that no. 487 was an exceptionally small mouse; it had a very short hind foot, only 22·0 mm., but, as has already been shown,

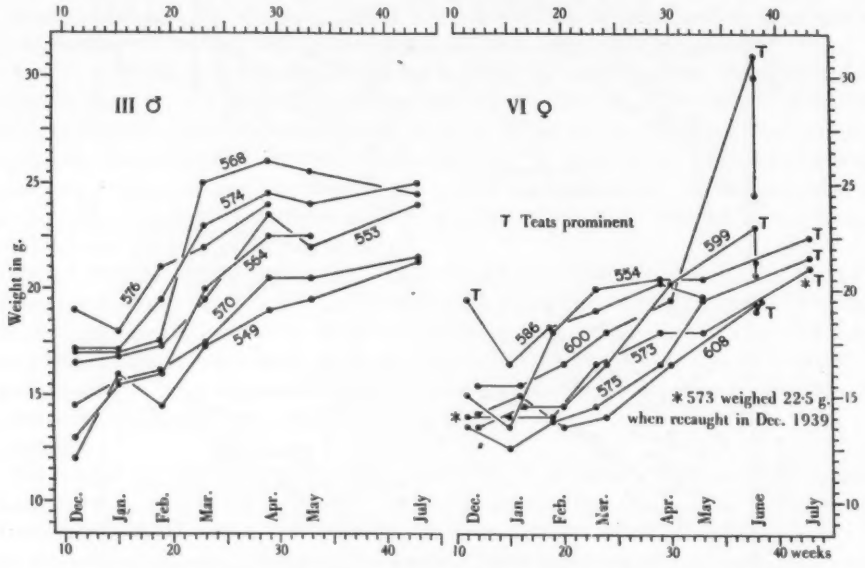


Fig. 8. Individual weight records of live *Apodemus* caught at monthly intervals in Holwood Park, December 1938 to July 1939.

there is not a very high correlation in adults between length of hind foot and size of mouse. That there is also no close association between foot length and rate of spring increase in weight can be seen in Fig. 9, where, however, one of the two dots in the bottom left-hand corner represents no. 487.

Fig. 8, group III, contains the weight graphs of seven male mice which were not caught until December. A comparison with Fig. 7, group I, suggests that nos. 564 and 549 were very young mice in November, as they showed a sharp rise in weight between December and January. The larger mice, nos. 553, 568 and 574, remained nearly stationary, and judging from other mice not included in the diagrams were likely to have had about the same November weight of 16·5–17·0 g., though there were also a few smaller mice in November which increased to that weight by December and January (Chart 1, D1, D2). The largest

mouse of all, no. 576, lost weight between December and January. They all showed the rapid increase which was common to male mice in the spring, and which slackened off sometimes at the end of March, sometimes in April.

Females. Fig. 7, groups IV and V and Fig. 8, group VI, give the weight graphs of a number of female mice. They show that young females soon fell behind young males of comparable November weight; no. 433 was an exception, keeping up with the males until after the February catch. There is much irregularity and overlapping among the females, and no general rule is apparent. Losses between December and January were frequent among both smaller and larger mice, while the three November adults, nos. 521, 425 and 459, still showed losses between January and March. Many even of the younger females showed no marked spring increase until after the March catch, though others started increasing between February and March, and a few between January and February. Between March and April all showed gains, some of which may have been due to early pregnancy; they were comparable in amount with those of the males between these months, and between January and February, but were not as great as those of the males between February and March.

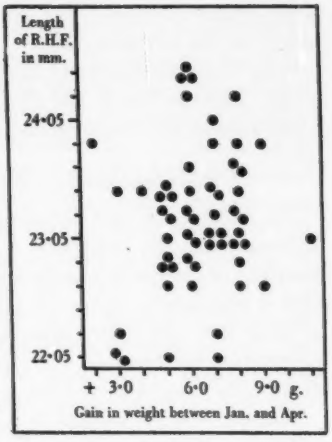


Fig. 9. Live male *Apodemus*, Holwood Park, 1939. To show that there is almost no correlation between length of hind foot and rate of spring increase in weight.

Advanced stages of pregnancy may be picked out by a sudden sharp rise in the graph, while a long vertical fall following this in the same week may be presumed to indicate the birth of young, though we had no other evidence of this when it occurred during release from captivity (see p. 150). Two examples may be taken from the graphs: no. 431, group IV and no. 600, group VI. No. 431 weighed 24·2 g. on 16 May. She was released near the same trap that afternoon, and found again next morning in another trap 10 yards off, together with three young which she had killed; her weight on this second day was only 18·6 g. She may have given birth to more young and eaten them, but in case there were still young unborn she was released again in the afternoon. Next day, 18 May, she was recaught in the original trap and then weighed only 17·4 g. No. 600 was caught on four consecutive days in June; she had the unusually heavy weight of 31·0 g. on the first day, had lost only a gram on the second day, but on the third day had dropped to 24·7 g., a weight which she retained on the fourth day. Apart from this drop in weight there was in this case nothing to indicate the birth of young.

Summary of information from individual weight graphs

Males. (i) Between November and December the smallest males gained most weight, larger mice remained stationary, while the largest lost weight.

(ii) Between December and January some of the smallest mice still gained a little weight, larger mice remained stationary or continued to lose weight.

(iii) Nearly all males put on weight after the January catch.

(iv) The period of most rapid male increase was between February and March.

(v) After March or April gains were smaller and some losses occurred.

(vi) Most of the older mice were still ahead of most of the younger mice at the beginning of the breeding season.

Females. (vii) Between November and December the weight changes of the females were similar to those of the males, except that the smallest females did not gain so much weight as the smallest males.

(viii) From December to March no general rule was apparent.

(ix) Between March and April increases were universal.

(x) In this year the breeding season started at the end of April.

(c) *Combined weight records. Spot diagrams of month to month weight changes*

The information just summarized comes from the very small sample of the population represented in the individual weight graphs. It is confirmed and amplified by the spot diagrams of month to month weight changes, which represent rather larger samples of the 1938–9 population. There are three series of these diagrams shown on Charts 1–3, and it is to these that the letters and numbers in the present section refer; each puts on record information not given by the others, and between them they enable most of the data from all the mice caught more than once to be utilized.

A series. Weight in each month plotted against loss or gain by the following month.*

C series. Weight in November plotted against loss or gain in each succeeding monthly interval.

D series. Weight in November plotted against weight in each succeeding month.

The CX and DX diagrams are given because we had a very large catch in January and the records of a number of new mice start then. The records of those males caught in November seem to show that autumn adults could still in January be distinguished by weight from autumn young. Judging from the data given in Chart 1, D2, the former might be expected not to have fallen to a weight of less than about 18·5 g. and the latter not to have increased to this weight. For adult females it is convenient to take a lower January limit of 17·5 g., for though some had undoubtedly dropped below this, it was then impossible to distinguish them by weight from growing young.

Adult males' winter loss of weight and early spring increase. The bulk of the November males of 18·5 g. and over lost weight by December, some of them as much as 2·5 g. (Chart 1, D1, C1).† Between December and January some again lost weight and some remained stationary (Chart 1, C2), the bulk of them in January being from 1·5 to 2·5 g. behind their November weight (Chart 1, D2). Between January and February some began to put on weight again, but some were stationary (Chart 1, C3, C3X) so that while a few were now ahead of their November weight, others were still from 1·5 to 2·5 g. behind it (Chart 1, D3). Between February and March nearly all showed big gains, most commonly of about 2·0 to 3·0 g. (Chart 2, C4, C4X); all now reached their November weight again, and some were as much as 3·5 g. ahead of it (Chart 2, D4). Between March and April the gains were very irregular and there were two losses, the range of increase being from −0·5 to +4·5 g. (Chart 2, C5). The bulk of the adult mice were now from 2·5 to 3·5 g. ahead of their November weight; the heavier the mouse in November the heavier it usually was in April, the range of weight being then from 23·5 to 30·0 g. (Chart 2, D5).

* A1 = C1; diagrams A2, 3 and 4 have been omitted for economy in space.

† The weights given in the text are the actual weights of the mice correct to 0·1 g. In the spot diagrams the data are grouped into 1·0 or 0·5 g. divisions, and so may sometimes appear not to correspond with the text.

Young males' winter and early spring changes. Of the November male young (under 17·5 g.), those under 15·5 g. had all gained weight by December, and, on the whole, the smaller they were the more they had gained; those under 11·5 g. had advanced very fast indeed, some gaining 5·0 to 6·0 g. and overtaking many of those above this weight which showed only small increases (Chart 1, D1). Between December and January the two mice which were smallest in November had both gained another gram, whereas the slightly larger mice had gained less than this or lost a little (Chart 1, C2); for the whole period, November to January, it remained true that the smaller the mouse in November the more it was likely to have gained (Chart 1, D2). Between January and February the bulk of the November young put on weight, some as much as 4·0 g. (Chart 1, C3), but the most rapid period of spring growth in this year, for the young as for the adults, was between February and March, when there were no losses in weight among the younger males but a wide range of increase, gains of 4·0 to 4·5 g. now being very common (Chart C4, C4X). Six weeks later, towards the end of April and just before the breeding season, some further big increases were shown, but on the whole the rate seems to have slowed down, the commonest gain being about 2·0 g. (Chart 2, C5, C5X); it still held good that the smaller a young mouse was in November the greater was its expected gain, the smaller young having by now overtaken the larger young within the range of 20·5–24·5 g. (Chart 2, D5).

An overlap in weight by February. Between the November young and the November adults an overlap in weight had already started in February, just before the period of maximum spring growth; this overlap was then between 18·5 and 20·5 g. (Chart 1, D3, D3X), while in March it was between 20·5 and 24·5 g. (Chart 2, D4, D4X), and in April between 21·5 and 26·5 g. (Chart 2, D5, D5X). From February on, therefore, there was no clear division by weight of the November male adults from the November male young, though the bulk of the former were still heavier than the bulk of the latter.

Male weight losses in the breeding season. After the start of the breeding season our trapping was confined to one only of our areas in May, to the other in June, and to the first again, but for two nights only, in July. This meant that we caught many less mice in each month, and among these there were very few of our November catch: in May only three of the November male adults and four of the young, in June two of the adults and nine of the young, in July one of the adults and one of the young.

In May and June the bulk of the November young weighed from 20·5 to 24·5 g., their average weight in May being 22·7 g., a gram higher than in June. The few remaining autumn adults were very scattered in weight, overlapping the young in range but with an average still a little ahead of them (Chart 3, D6X, D7X). Between April and May there were nearly half as many losses as gains, and although these had apparently no relation to the weight of the mouse in November (Chart 3, C6X), the losses were confined to mice which in April were 21·5 g. and over, while the laggards then under 21·0 g. all showed gains (Chart 3, A6); in all, nine mice lost weight, nineteen gained weight, and four remained the same. Between April and June losses outnumbered gains by fifteen to twelve (Chart 3, A7). It is perhaps significant that all these losses occurred in the early part of the breeding season, during the two months in which pregnant females first appeared in the traps. The net result, a male population almost stationary in weight, is best seen in the weight histograms (Fig. 10).

A possible summer period of further weight increase. Of the thirty-four males caught in May only thirteen were recaught during the two nights when we trapped over the same

area in July. This is a very small number from which to draw conclusions, but since eleven of the thirteen had gained from 1·0 to 2·5 g., and only two had lost weight, it is possible that a new period of weight increase had followed the period of losses (Chart 3, A 8). The increase in weight of this very small sample of the July population is again best seen in Fig. 10; leaving out the new season's young, the average weight was 24·3 g., whereas the May and June averages were only 22·9 and 22·8 g.* It may be noted here that no young were born in the traps or cages in July, and that none of the females caught appeared to be in an advanced state of pregnancy, the heaviest weighing only 22·3 g.

Table 3. *Adult females' loss of weight after capture*, 1938–9

	424	425	422	415	416		488	495	476	465	516	466	521
14 Nov.	—	—	23·7	25·7[1]	30·6[2]	21 Nov.	—	—	21·5	22·9	—	24·8	—
15	16·6	20·0	—	—	—	22	17·6	20·6	21·7	21·8	—	—[3]	—
16	14·2	18·5	22·1	—	—	23	18·8	21·5	—	21·9	23·0	—	—
17	16·0	—	18·3[1]	—	—	24	15·9	19·2	18·8	19·9	22·1	—	24·8
	—	—	—	—	—	25	16·8	19·4	20·6	21·3	19·9	—	21·6
	—	—	23·0	18·3	26·5		—	—	—	—	—	18·6	—

	715	743	452	620	431	717		768	643	599	637	660	600
15 May	—	—	—	23·2[4]	—	—	19 June	22·3	23·0	23·1	24·7	26·5	31·1[7]
16	20·0	20·9	22·2	—	24·2	26·3	20	19·7	21·3	21·6	22·4	20·0	30·0
17	19·8	20·5	20·5	17·6	18·6[5]	24·5	21	—	19·9	21·1	19·4	—	24·7
18	19·2	18·3	18·6	—	17·4	19·5[6]	22	—	20·2	20·3	—	—	24·7

[1] 422. Ill, kept in for two days; weight increased again to 23·0 g.

[2] 415 and 416. Both appeared nearly parturient and were not released. Young subsequently heard in cage, but died and were eaten. Mothers weighed again on 24 November and released.

[3] 466. Vaginal discharge on 22 November; not weighed, kept in cage. 24 and 26 November, young heard. 1 December, six dead young seen, mother weighed 18·6 g.

[4] 620. Appeared nearly parturient, kept in. Young seen and heard next day; some killed and partly eaten by mother, the rest placed by us in some hay with the mother in the woods where she had been caught. Young, all dead, found still in hay on third day; mother recaught weighing 17·6 g.

[5] 431. Three dead young in trap.

[6] 717. Bleeding at vulva.

[7] 600. This very heavy mouse had a very long right hind foot.

The remaining weight losses all occurred while the mice were set free overnight.

Adult females in winter. Among the females the November adults were not so clearly marked off by weight from the November young as among the males. Teats were prominent in November in all female mice of 17·5 g. and over, and also in two out of three of about 16·5 g. A heavy female rarely dropped in weight to less than 17·5 g. when known or supposed to have given birth to a family (p. 150, and Table 3); this is therefore a convenient weight above which to regard all females as adult, and 'November adults' of 1938–9 will here mean those shown on the graphs as 17·5 g. and over in November, though it is probable that females of a lower weight than this do become pregnant.

* The averages given on p. 155 were for the November young only.

Late autumn pregnancies. Among these adults the biggest losses between November and December were in two mice which fell from 24·8 to 21·5 and to 17·5 g. respectively (Chart 1, D1). The weight graph of one of these has been seen in Fig. 7, group IV (no. 521). Most of the loss had already occurred when the mice were weighed for a second time in November (see Table 3)—the first mouse (no. 521) when recaught after a night's release from captivity, the second (no. 466) at the end of a week's captivity after giving birth to a family of at least six young. Such losses are comparable with those in May when it was known or suspected that pregnant females had given birth to young (p. 158). Three females of about 20·0 g. in November lost from 1·5 o 2·5 g. by December, while one of 17·5 g. remained at this weight (Chart 1, C1, D1); these mice may have been nursing mothers in November (p. 161).

Among females, no conspicuous weight increase until after March. Between December and January one mouse in this group gained weight (but not as much as it had previously lost), three lost weight, and two remained the same (Chart 1, C2); between January and February one gained a little, two lost weight and two remained the same (Chart 1, C3, C3X); between February and March, the most rapid period of male growth, three lost weight and two remained the same (Chart 2, C4, C4X). The net result was that by March the five surviving November adults were all from 2·0 to 5·0 g. behind their November weight, in great contrast to the adult males which had all by now overtaken that weight, some showing gains of up to 3·5 g. (Chart 2, D4). In the six weeks between the March and April trappings a period of growth at last set in for all females, and the five surviving adults showed increases of up to 4·0 g. (Chart 2, C5, C5X), all but one of them having now nearly caught up with their November weight (Chart 2, D5). After April only one of these adults was ever caught again—in June, when its teats were prominent and its weight was 26·2 g.

Young females compared with young males in winter and early spring. Among the November young only three very small females of 8·0, 8·1 and 8·7 g. in November put on much weight between November and December, and of these only the largest put on as much as the males of similar size (Chart 1, C1, D1). The individual graphs of these three were given in Fig. 7, groups IV and V (nos. 431, 450 and 433). Young females of 10·5–13·0 g. in November did increase slightly, but not nearly as much as the males of this weight, while from 13·5 to 17·5 g. there were small changes in both directions. Between December and January the very young mice ceased their rapid increase, two of them actually losing weight, while those of intermediate size showed some gains and some losses (Chart 1, C2); young females behaved very like young males in this month, the juvenile population as a whole remaining stationary (Chart 1, C2). Between January and February the very youngest females were once more putting on weight, but from 10·5 g. upwards losses and gains about balanced one another; this was in contrast to the males, only two of which showed loss in weight, while there was a high proportion of big increases (Chart 1, C3, C3X). Between February and March the contrast was still greater; there were fewer losses among the young females but no very big increases, and four remained the same (Chart 2, C4, C4X). The net result was that the mice which were smaller in November had gained from 5·0 to 9·0 g. by March while the rest showed intermediate gains and losses in rough accordance with their November weight (Chart 2, D4). Already since February some of the November young had overtaken some of the November adults within the range of 16·5–18·5 g. (Chart 1, D3X and Chart 2, D4X).

The six weeks between the commencement of the March and the April trappings were, as stated, a period of growth for all females. Young mice as well as old showed gains of from 0·5 to 4·0 g., about 2·0 g. being the commonest (Chart 3, A 5 and Chart 2, C5, C5X). These gains were much the same as those of the males for this period, since the male increase had by now slowed down, but nevertheless the bulk of the females were still far behind males of similar age (Chart 2, D5, D5X) and, in spite of further male losses in May and June, never overtook them unless in advanced pregnancy (Chart 3, D6X, D7X).

Spring pregnancies. As already seen in November, advanced pregnancy reveals itself in these diagrams by a wide gap between an individual and the bulk of the female population,* just as it is disclosed in the individual weight graph by a sudden sharp rise (p. 153, and see also Table 3, p. 156). There were already two such individuals in April (Chart 2, C5, C5X), mice which had weighed only 15·2 and 16·2 g. in November (Fig. 7, group V, nos. 497 and 457). In May four other females were widely marked off from the rest by exceptionally big increases of from 8·5 to 11·0 g. (Chart 3, A 6); the individual graphs of two of these are seen in Fig. 7, group IV (nos. 431 and 419). All four had prominent teats, and two gave birth to young before it was possible to release them; a third (no. 717), after being released twice, was found to be bleeding at the vulva when caught for a third time, and as she had dropped in weight from 26·3 to 24·5 to 19·5 g., may be presumed to have given birth to a family during her second night of freedom. In June there were two other exceptional increases, of 11·5 and 8·5 g. (Chart 3, A 7); after release and recapture the weight of these mice was found to have dropped from 31·1 to 24·7 g. (no. 600), and from 26·5 to 20·0 g. (no. 660), so that they also may be supposed to have given birth to young in the interval; the individual graph of the first of these two is shown in Fig. 8, group VI (no. 600, see also p. 153).† The next largest increase between April and June was of 5·5 g., but in this mouse the drop in weight on recapture was gradual and not easy to interpret; it was from 23·0 to 21·3 g. after the first night, and from this to about 20·0 g. after the second and third nights (Table 3, p. 156, no. 643).

In June and July prominent teats were noted in all but two of the thirty-one females weighing 17·5 g. and over, whereas in May, out of twenty-five such mice, they were noted in only four besides those four already mentioned as in advanced pregnancy.

Apart from these cases of pregnancy, the bulk of the young females showed advances of from 1·5 to 3·5 g. between both April and May, and April and June; two remained the same in May and two had lost weight (Chart 3, C6X, C7X). The only female which lost weight between the April and June catching was one of the two which had put on so much weight between March and April as to have suggested an early pregnancy (see above, and Fig. 7, group V, no. 497).

None of the females caught as adults in November, or presumed to have been adults

* It must not be forgotten that any *single* exceptional measurement might be due to an exceptionally full stomach (p. 144 and Fig. 4) or to an error in recording. Single errors may also creep in in some such way as follows. In March 1941 a female mouse of 29·2 g. was recorded on capture as appearing to be in an advanced state of pregnancy. She was not released and next day was found dead in the cage, and on dissection proved to have a mass of tapeworm cysts attached to the small remaining fragments of the liver, and occupying most of the greatly distended abdomen. This mass weighed 13·1 g., which, subtracted from the apparent weight of the female, would leave her true weight, without liver, at 16·1 g., that of a not yet fully grown mouse; when first caught in December she had the adult weight of 19·5 g. Had she not died she would have remained in the records as a unique case of pregnancy in the middle of March.

† This mouse, no. 600, was apparently an exceptionally large individual. The length of its right hind foot was bout 24·0 mm., much above that of the average female.

then by their weight in January, were recaught in May or July, and only two* in June (Chart 3, D7X).

(d) Combined weight records. Monthly histograms

In Figs. 10 and 11 are frequency histograms† showing the monthly weight distribution of the Holwood Park 1938–9 male and female populations. Use is here made of all the mice caught, even those caught only once. In Fig. 12 these histograms are used as a background to a similar series for 1937–8. In that year, instead of an intensive fortnight's trapping at

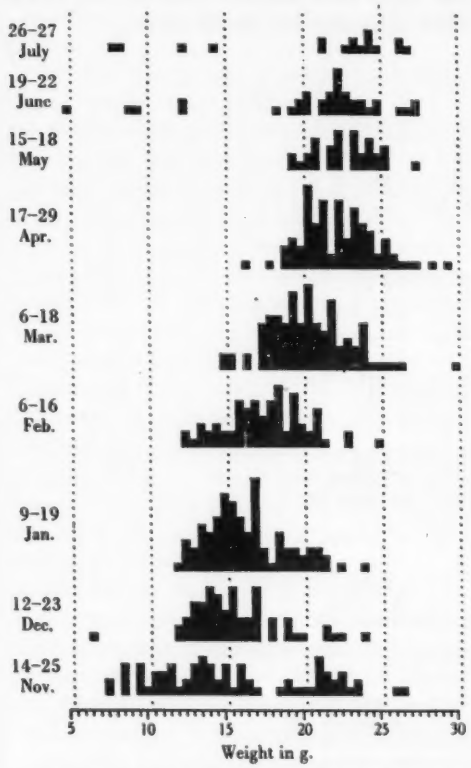


Fig. 10. Monthly weight frequencies of live male *Apodemus*, Holwood Park, 1938–9.

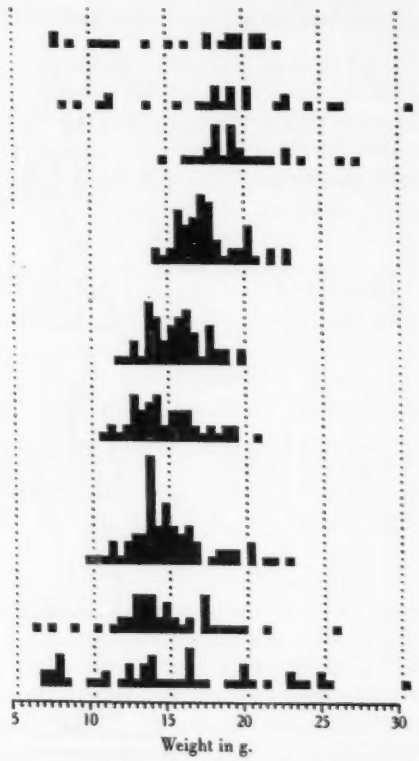


Fig. 11. Monthly weight frequencies of live female *Apodemus*, Holwood Park, 1938–9.

4-weekly intervals, each month's catch was spread over the whole month; we trapped each week, covered a wider area, and shifted our trapping sites according to a different plan. Individual mice were caught at much less regular intervals and so gave less satisfactory records, while the monthly grouping is not strictly comparable with that for 1938–9. The general picture of weight increase in the two years is the same, however, and presents in another way information already given by the individual weight graphs and the month

* One, no. 637, was not caught in April, so is not included in Chart 3, C7X.

† In these diagrams and in Fig. 12 the unit of frequency on the vertical scale equals the distance between the dots on the vertical lines.

to month spot diagrams: a late autumn population reducing its wide weight range by the growth of the small mice and the weight losses of the large mice; a nearly stationary mid-winter population; and a rapid early spring increase among the males followed by a later, lesser increase among the females. Reduced trapping during the breeding season of 1939 extended the picture for that year, and showed a slowing down of weight increase and the reappearance of a wide weight range. Comparison of Figs. 10 and 11 shows once more that in 1938–9 the female population kept up with the male until after January; between February and March was the period of most rapid male increase, whereas the female population did not as a whole gain much weight until after the March catch.

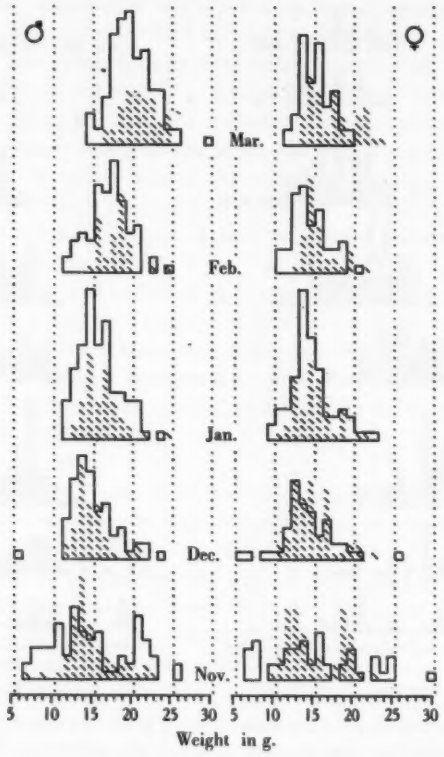


Fig. 12. Monthly weight frequencies of live *Apodemus*, Holwood Park. Plain outline, 1938–9. Hatched, 1937–8.

Where differences between the two years can be detected they can probably be attributed to the marked weather differences. The more obvious differences in the histograms are as follows:

(i) More young of both sexes in November 1938–9.

(ii) More males of over 20·0 g. and females of over 22·0 g. in November 1938–9.

(iii) A group of smaller mice of both sexes in February 1938–9, not present in 1937–8.

(iv) A group of females of over 20·0 g. in March 1937–8, not present in 1938–9.

With regard to these four points, the following comments may be made:

(i) The autumn of 1938 was exceptionally mild and breeding continued late. A number of 7·0–10·0 g. youngsters were caught in November and some even in December (cf. p. 150). All the females of over 17·0 g. and two of 16·5 g. had prominent teats; some were probably nursing young in the nest, while others may have been pregnant (cf. pp. 156–7). All these females were set free each day but they were often caught three nights running and we are afraid that some of the nest young may have died meanwhile (cf. p. 150 and footnote); if so, this would have reduced the number of 7·0 to 10·0 g. youngsters in the December population.

(ii) The population of November 1938 differed from that of November 1937, not only because of the number of very young mice but also because of a greater number of heavy adults. The male group of over 20·0 g. is especially conspicuous in the histogram of Fig. 10, and in weight distribution resembles on a small scale the total male population in April, the beginning of the 1939 breeding season. Most of this group, like the similar group of females, may still have been breeding, but more information is needed about the connexion between weight, season and fecundity in *Apodemus*, which we believe has already been collected by Brambell and Rowlands (pp. 146–7). In the November 1937 population there were far fewer males of over 18·0 g. in proportion to the intermediate group of 13·0–18·0 g. The earlier onset of cold weather is likely to have checked breeding, and the consequent recession of the male reproductive organs to have been accompanied by a loss in weight. The November catch in this year indeed appears comparable with that of December 1938, when adult males had lost from anything up to 2·6 g. of their November weight (p. 154 and Chart 1, C1, D1).

As regards adult females, it is true that in November 1937 we caught at least ten which still had prominent teats, but though eight of these weighed over 18·0 g. and are part of a conspicuous group at the tail of the histogram, there were none comparable with the 1938 group of over 22·0 g. Like the rarity of youngsters of under 10·0 g., and the evidence from the adult males, this again points to the 1937 breeding season having ended before November. One unique little mouse must not be forgotten however: a male of only 7·2 g. caught on 2 December of that year.

(iii) The next conspicuous difference in the histograms is in February. While we can be fairly certain that the differences in November are connected with the long mild autumn of 1938, the reason for this February difference is less clear. Since in December 1938 we caught eight females of over 16·0 g., and among these were five new mice still with prominent teats, it is tempting to think that these or similar late breeders may have been responsible for the batch of small mice characteristic of February 1939 and not found in February 1938. Individual analysis, however, shows that this conspicuous group is as likely to be composed of mice whose spring growth was for some reason delayed, as to be the result of autumn breeding prolonged into December (see also Chart 2, C4, C4X).

(iv) The only other conspicuous difference between the two series of histograms is in March, where among the females of 1938 there is a prominent group of over 20·0 g., absent in 1939. In March 1938 there were three almost unbroken weeks of warm, sunny weather, and towards the end of the month two families were born: one in a cage on 25 March after the mother had been captive for four days, and one in a trap on 26 March. On this we stopped trapping, hoping that we had not already interfered appreciably with the summer population by repeatedly catching and keeping in captivity members of the winter population which might otherwise have started to breed. In 1939, March was a comparatively cold

month, and early April very cold, so that it was not until the beginning of May that we began to get births in the traps and cages (cf. p. 158).

5. Summary

1. Introduction (p. 136).

2. A discussion of the factors which limit the size of the samples caught when trapping *Apodemus* populations for comparison by statistical methods (p. 137).

3. A discussion of the value and reliability of certain conventional measurements, (*a*) in determining the probability that an individual mouse is fully grown, (*b*) in comparing fully grown populations from different localities (pp. 139–44).

The part played by the stomach content (pp. 144–5) and by the reproductive organs (pp. 145–9) in the varying weight of a mouse (p. 141).

4. (*a*) An analysis of monthly weight records of a living population in a single locality (pp. 149–50). These records were made in order to determine the best time of year to trap when comparing populations from different localities; but they also give a picture of the growth of *Apodemus* and how it is influenced by season. The picture is constructed from:

(*b*) Graphs of the weight records of those individual mice caught most frequently in the trapping season 1938–9 (pp. 151–4).

(*c*) Spot diagrams in which the weights, or weight changes, of individual mice are plotted against their weights in a previous month (pp. 154–8 and Charts 1–3).

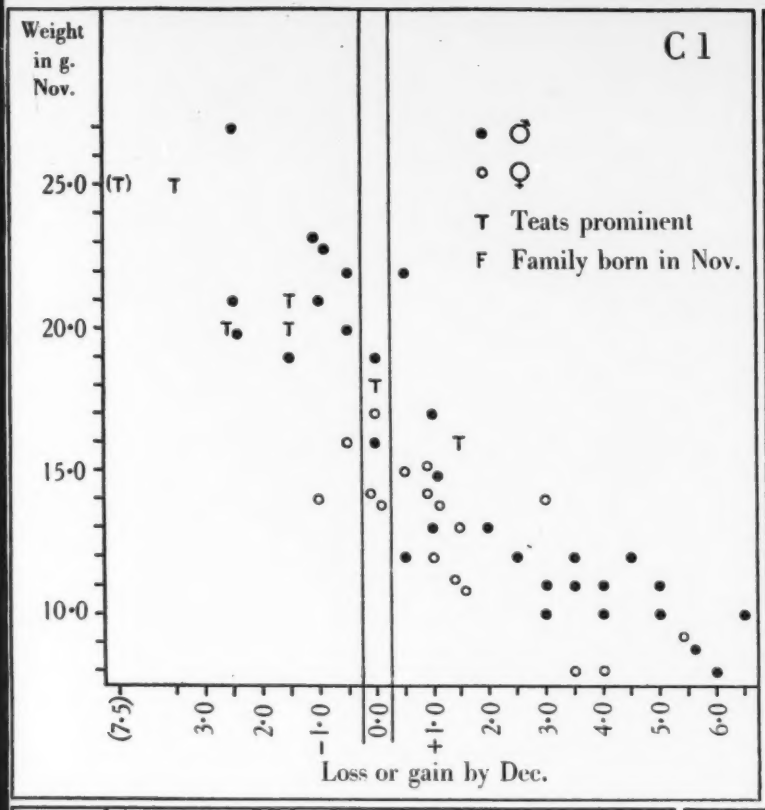
(*d*) Histograms comparing the weight frequencies of the sample populations caught in each month of the trapping season 1938–9; and a comparison of these histograms with a similar series for 1937–8 when weather conditions were markedly different (pp. 159–61).

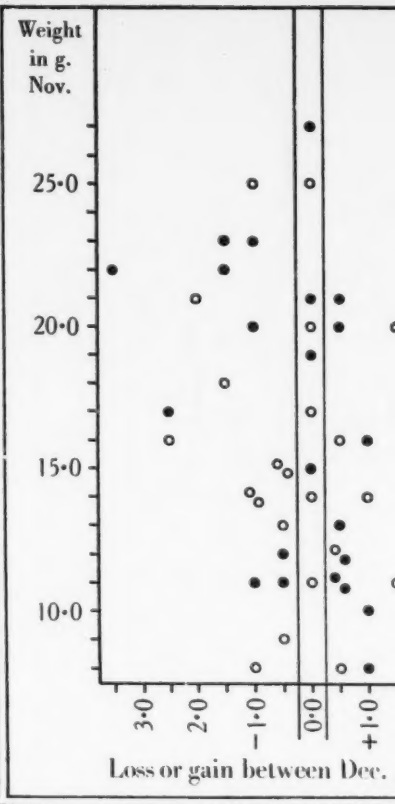
Our grateful thanks are due to Miss Enid Harris for her dextrous help in field and laboratory, to Miss Joyce Townend for her skill and care in making the diagrams, to Prof. E. S. Pearson for a critical reading of the manuscript which led to many improvements, and to Dr G. M. Morant and Mrs Karl Pearson for useful suggestions when reading the proofs.

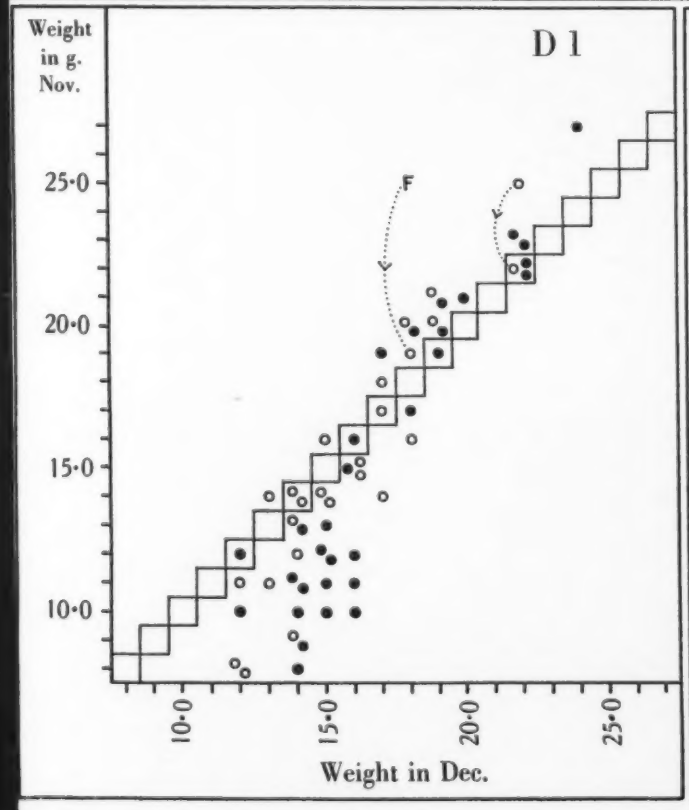
REFERENCES

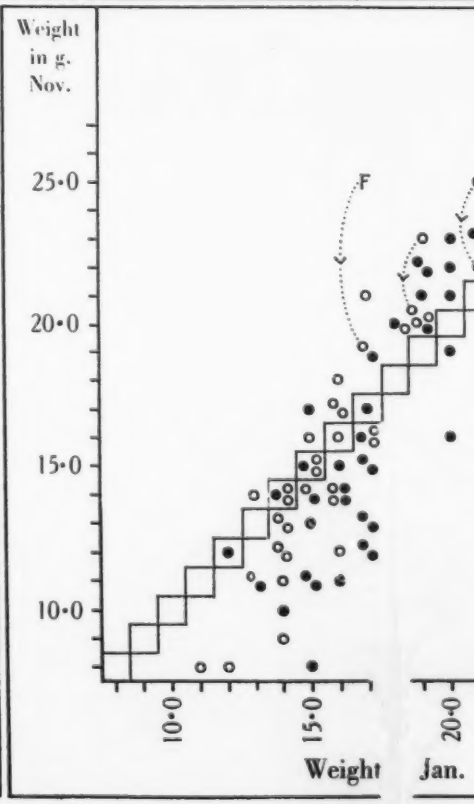
Brambell, F. W. R. & Rowlands, I. W. (1936). *Philos. Trans.* B, **226**, 71.
Burt, W. H. (1940). *Misc. Publ. Mus. Zool. Univ. Michigan*, no. 45.
Chitty, D. (1937). *J. Anim. Ecol.* **6**, 36.
Elderton, E. M. & Moul, M. (1928). *Ann. Eugen., Lond.*, **3**, 277.
Elton, C. E., Ford, E. B. & Baker, J. R. (1931). *Proc. Zool. Soc., Lond.*, 657.
Evans, F. C. (1942). *J. Anim. Ecol.* **11**, 182.*
Ruger, H. A. (1933). *Ann. Eugen., Lond.*, **5**, 59.
Ruger, H. A. & Stoessiger, B. (1927). *Ann. Eugen., Lond.*, **2**, 76.
Sumner, F. B. (1926). *J. Mammal.* **7**, 149.
Sumner, F. B. (1927). *J. Mammal.* **8**, 59.

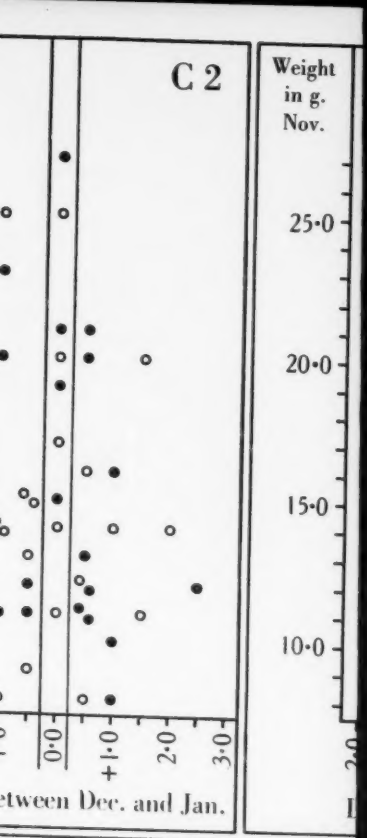
* Received after the present paper had gone to press.

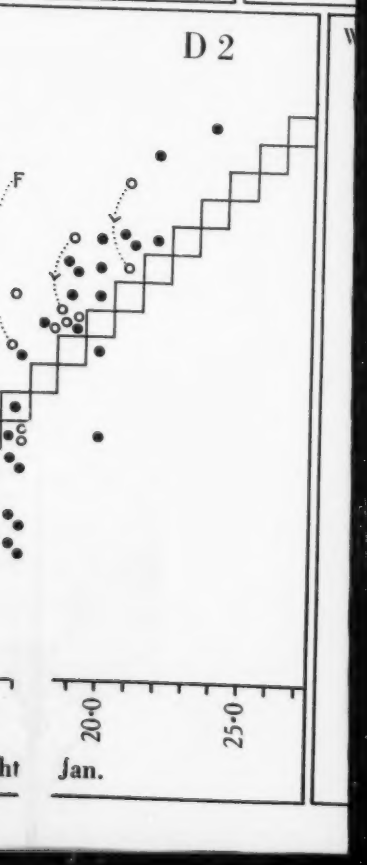

C 1
Weight in g. Nov.
♂
♀
T Teats prominent
F Family born in Nov.
25·0
20·0
15·0
10·0
(T)
T
(7·5)
3·0
2·0
−1·0
0·0
+1·0
2·0
3·0
4·0
5·0
6·0
Loss or gain by Dec.

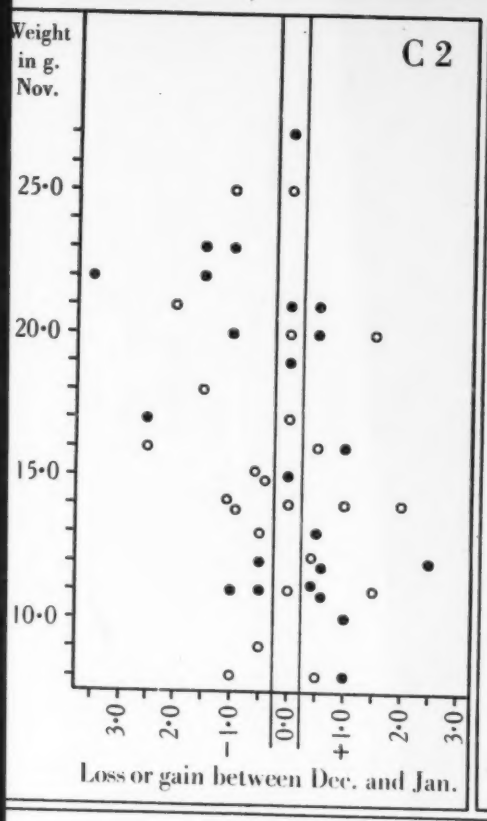

Weight in g. Nov.
25·0
20·0
15·0
10·0
3·0
2·0
−1·0
0·0
+1·0
Loss or gain between Dec.

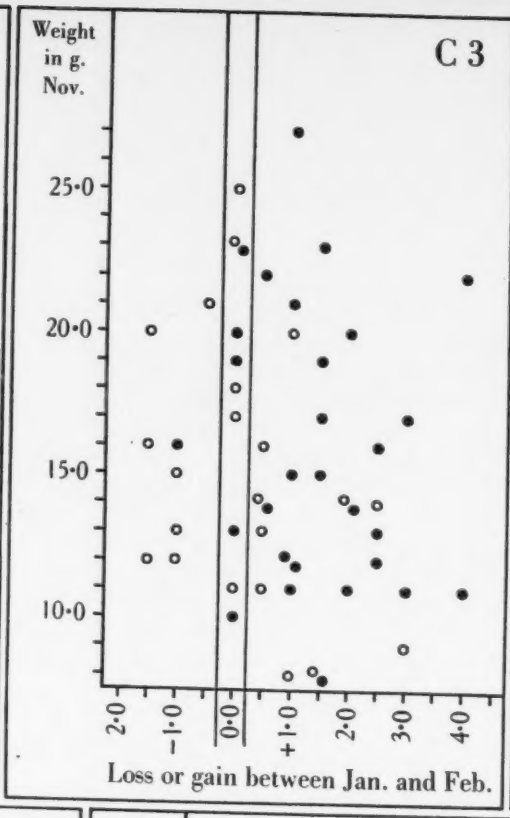

D 1
Weight in g. Nov.
F
25·0
20·0
15·0
10·0
10·0
15·0
20·0
25·0
Weight in Dec.

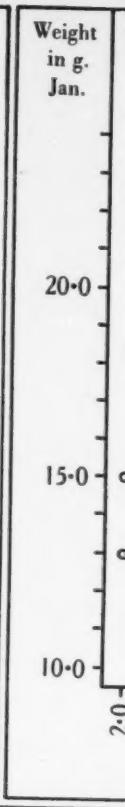

Weight in g. Nov.
F
25·0
20·0
15·0
10·0
10·0
15·0
20·0
Weight Jan.

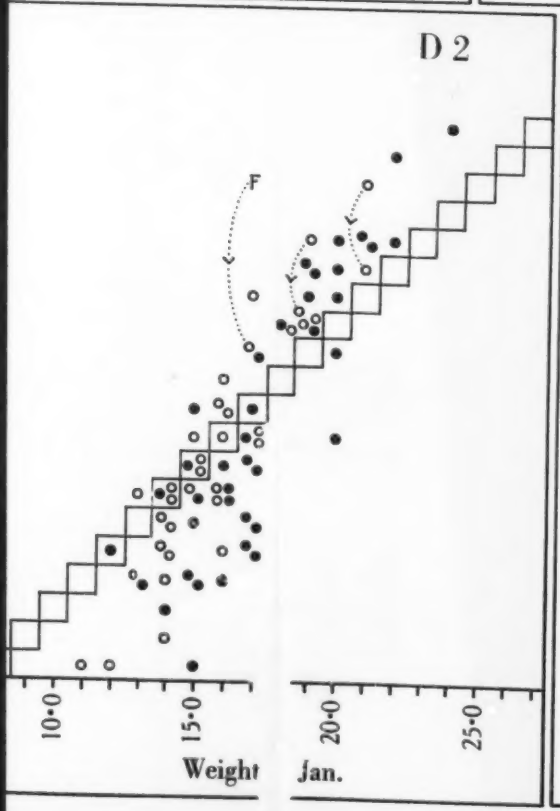

C 2
Weight in g. Nov.
25·0
20·0
15·0
10·0
0·0
+1·0
2·0
3·0
etween Dec. and Jan.


D 2
F
20·0
25·0
ht
Jan.


C 2
Weight in g. Nov.
25·0
20·0
15·0
10·0
3·0
2·0
– 1·0
0·0
+ 1·0
2·0
3·0
Loss or gain between Dec. and Jan.


C 3
Weight in g. Nov.
25·0
20·0
15·0
10·0
2·0
– 1·0
0·0
+ 1·0
2·0
3·0
4·0
Loss or gain between Jan. and Feb.


Weight in g. Jan.
20·0
15·0
10·0


D 2
F
10·0
15·0
20·0
25·0
Weight
Jan.

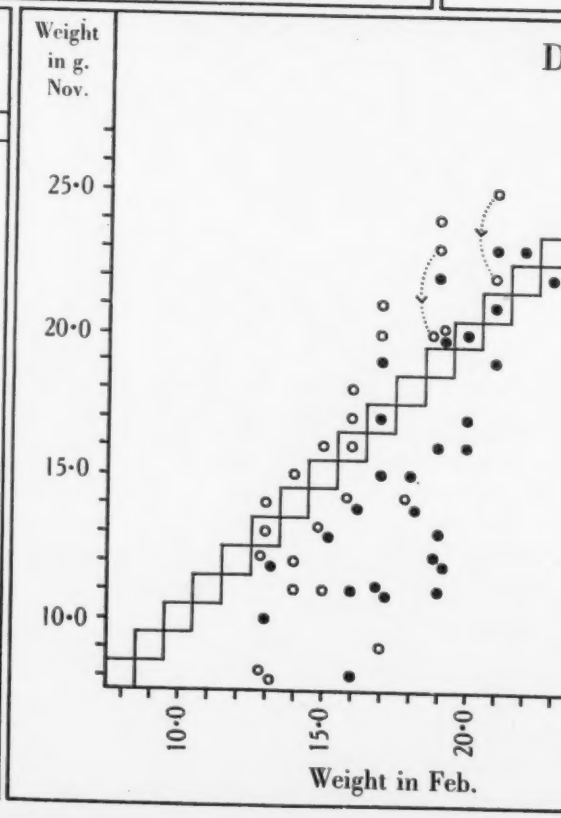
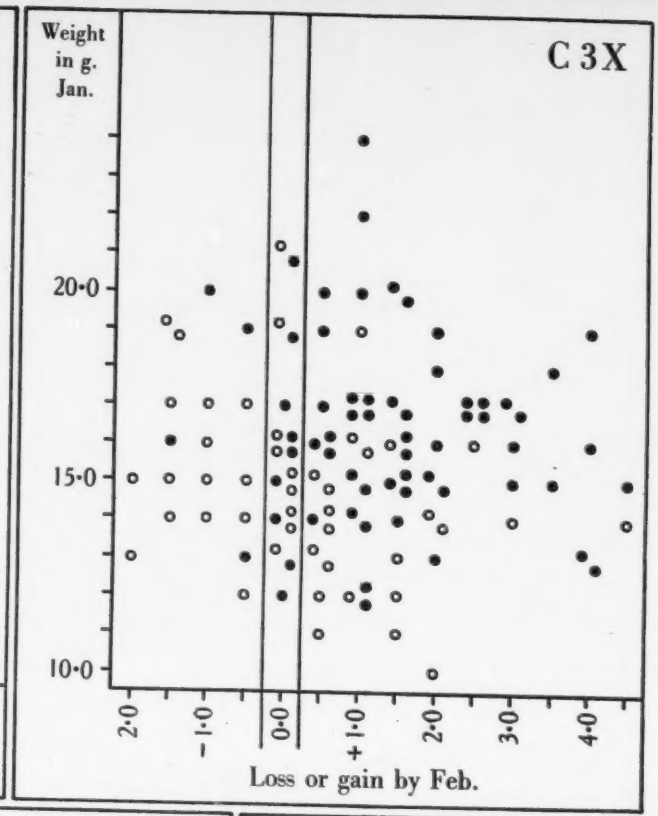
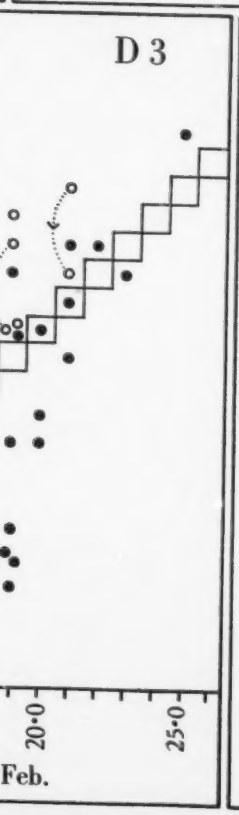
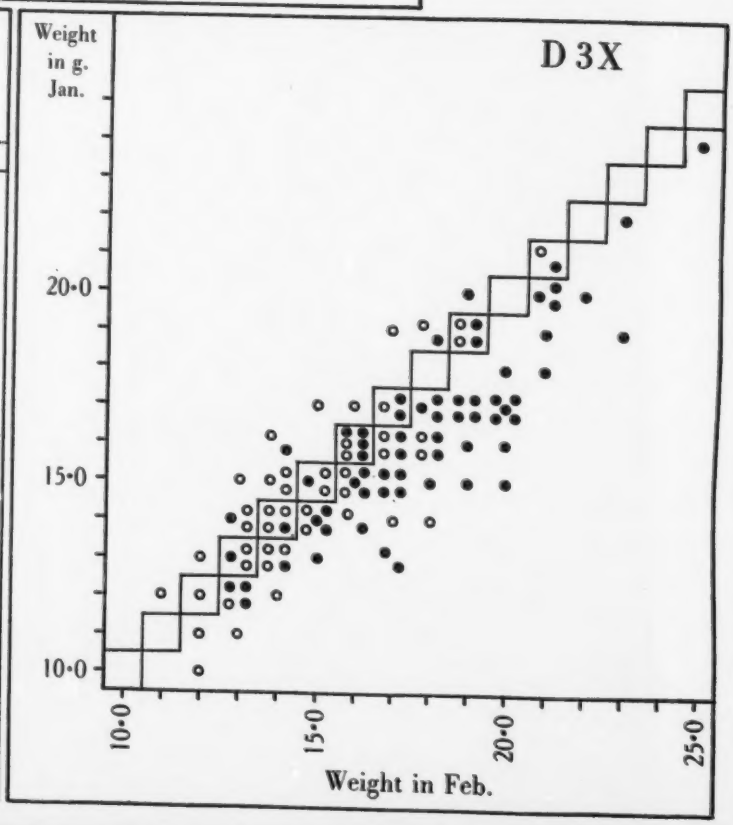
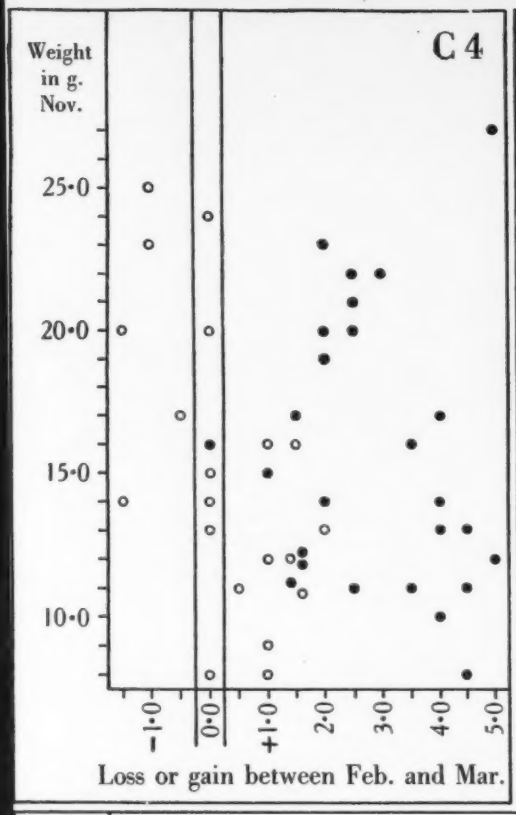

Weight in g. Nov.
25·0
20·0
15·0
10·0
10·0
15·0
20·0
Weight in Feb.

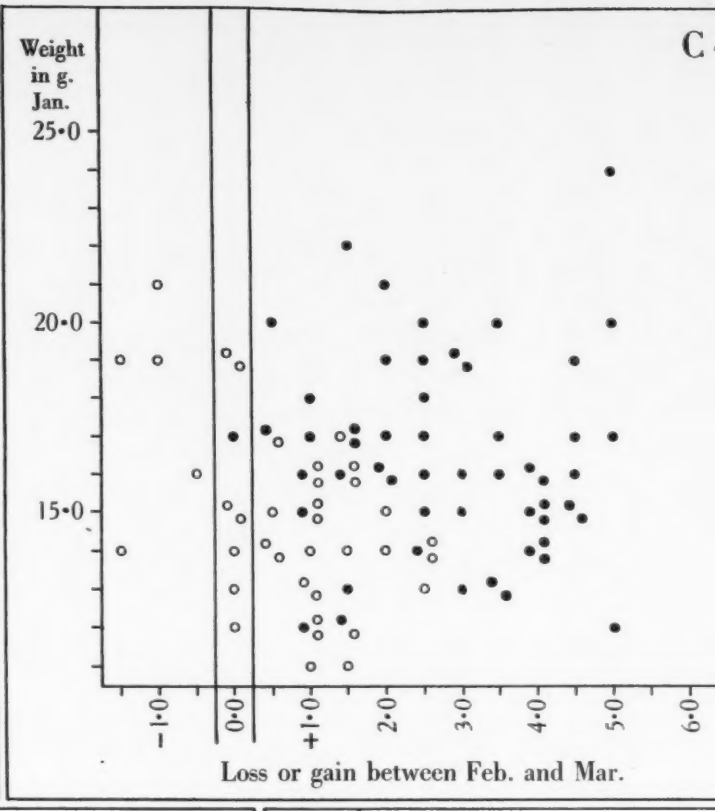
CHART 1

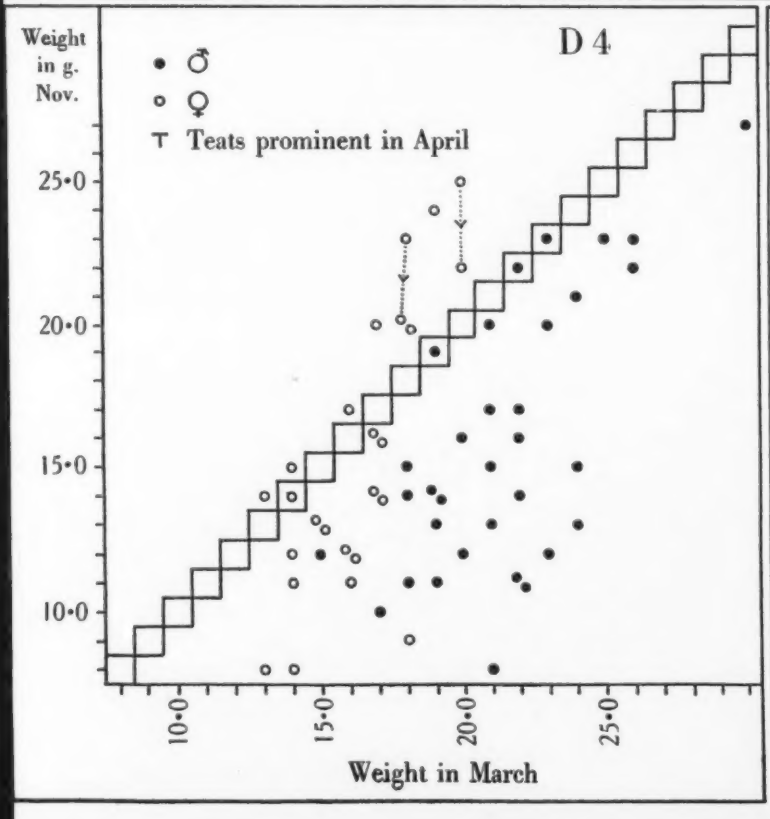

C 3X
Weight in g. Jan.
20·0
15·0
10·0
2·0
− 1·0
0·0
+ 1·0
2·0
3·0
4·0
Loss or gain by Feb.

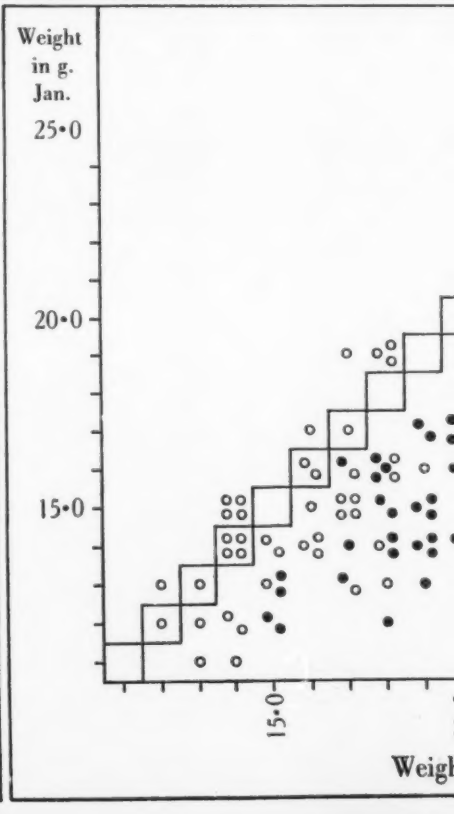

D 3
20·0
25·0
Feb.

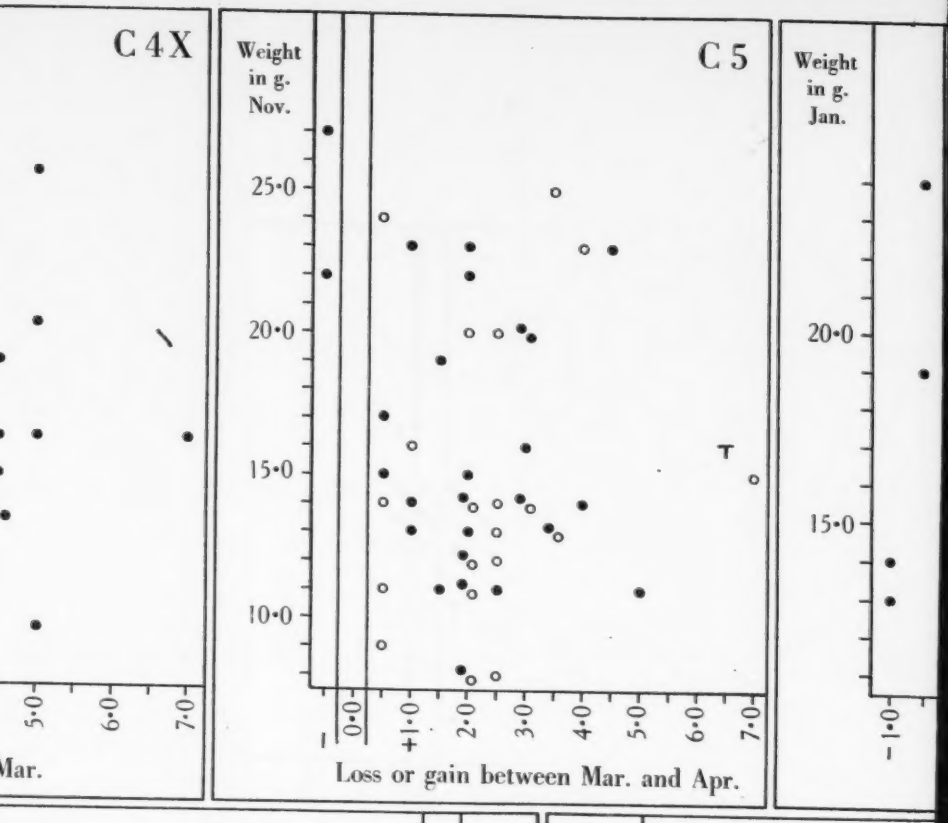

D 3X
Weight in g. Jan.
20·0
15·0
10·0
10·0
15·0
20·0
25·0
Weight in Feb.

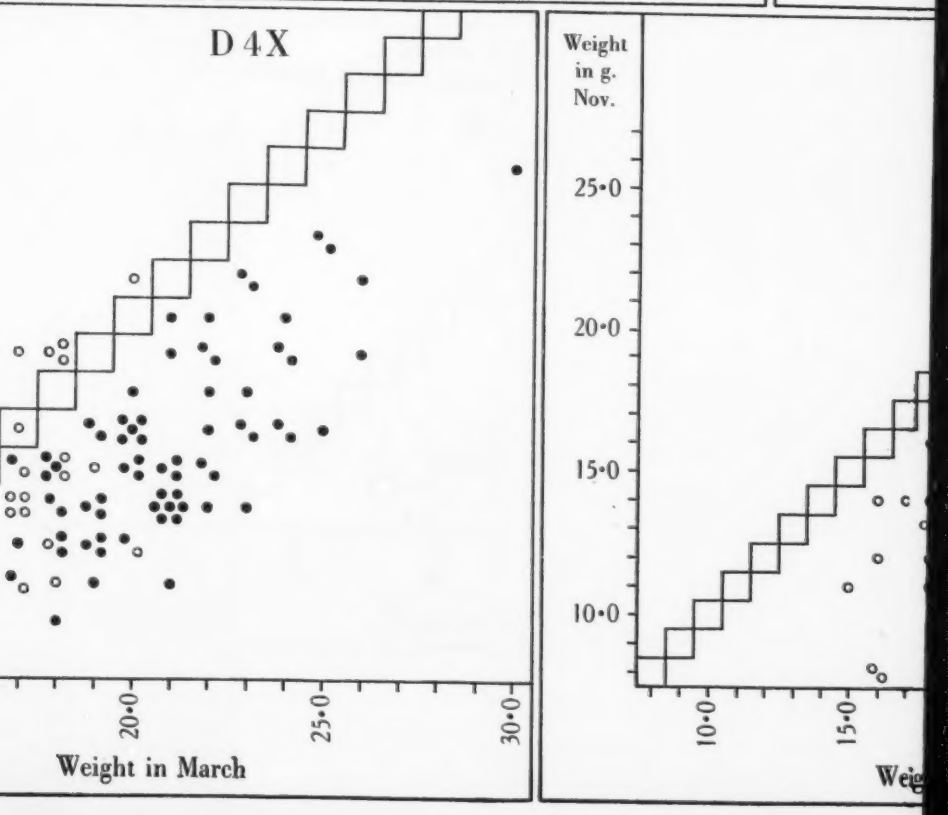

C 4
Weight in g. Nov.
25·0
20·0
15·0
10·0
−1·0
0·0
+1·0
2·0
3·0
4·0
5·0
Loss or gain between Feb. and Mar.

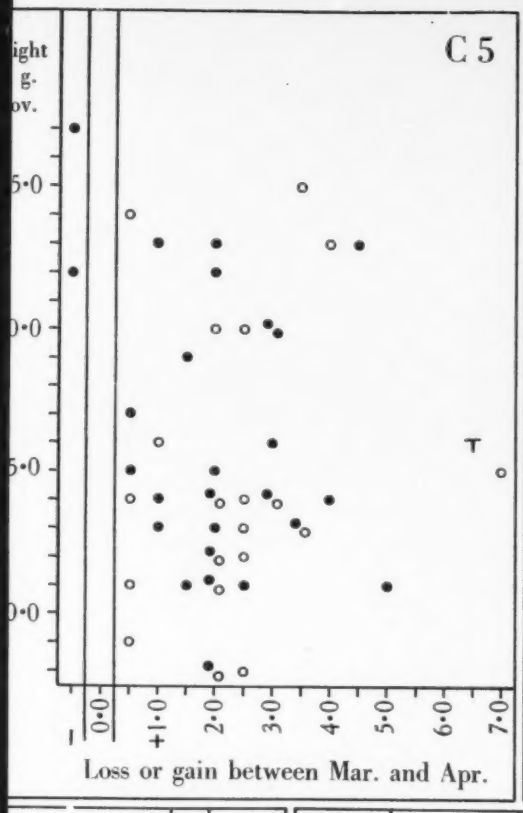

Weight in g. Jan.
25·0
20·0
15·0
−1·0
0·0
+1·0
2·0
3·0
4·0
5·0
6·0
Loss or gain between Feb. and Mar.

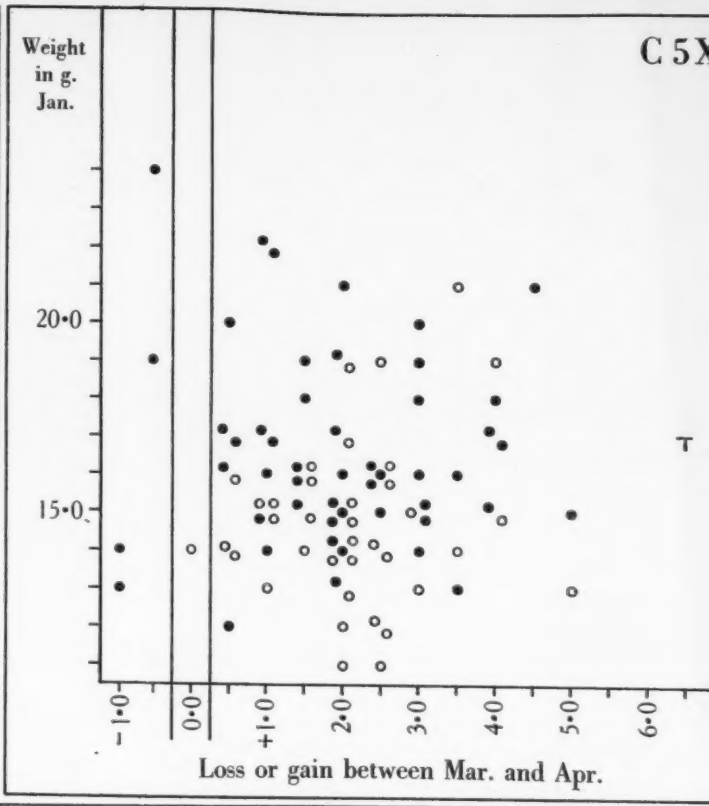

D 4
Weight in g. Nov.
♂
♀
Teats prominent in April
25·0
20·0
15·0
10·0
10·0
15·0
20·0
25·0
Weight in March

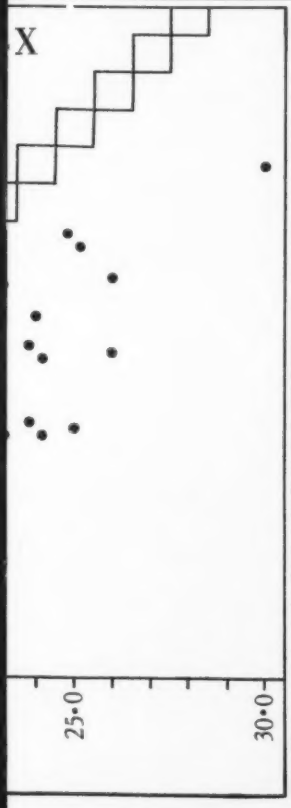

Weight in g. Jan.
25·0
20·0
15·0
15·0

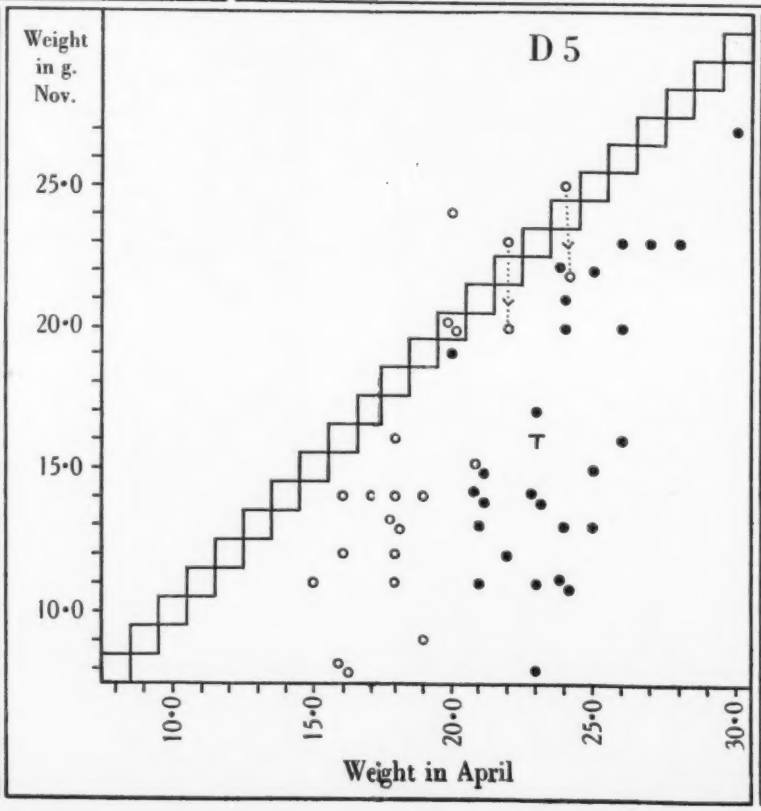

C 4X
Mar.
5·0
6·0
7·0
Weight in g. Nov.
25·0
20·0
15·0
10·0
C 5
−
0·0
+1·0
2·0
3·0
4·0
5·0
6·0
7·0
Loss or gain between Mar. and Apr.
Weight in g. Jan.
20·0
15·0
− 1·0

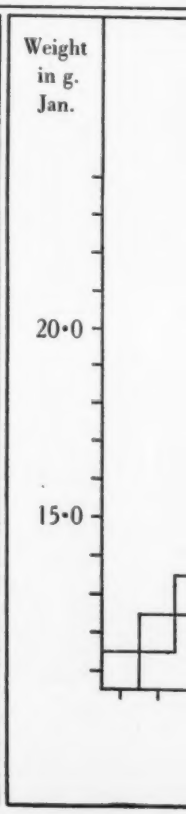

D 4X
20·0
25·0
30·0
Weight in March
Weight in g. Nov.
25·0
20·0
15·0
10·0
10·0
15·0

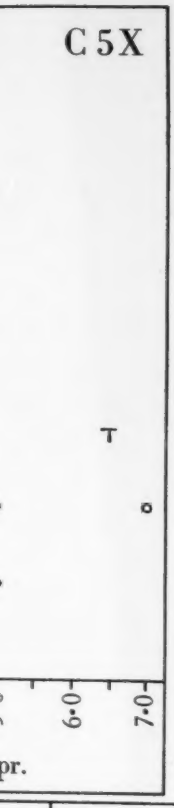

C 5
ight
g.
ov.
5·0
0·0
5·0
0·0
– 0·0 +1·0 2·0 3·0 4·0 5·0 6·0 7·0
Loss or gain between Mar. and Apr.

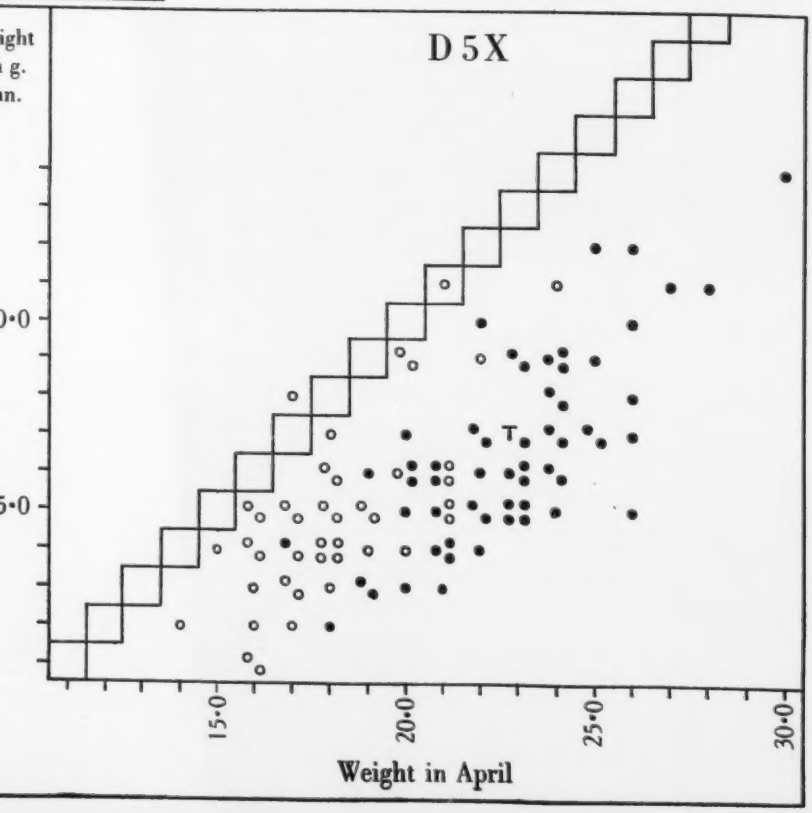

C 5X
Weight
in g.
Jan.
20·0
15·0
– 1·0 0·0 +1·0 2·0 3·0 4·0 5·0 6·0
Loss or gain between Mar. and Apr.


X
25·0
30·0

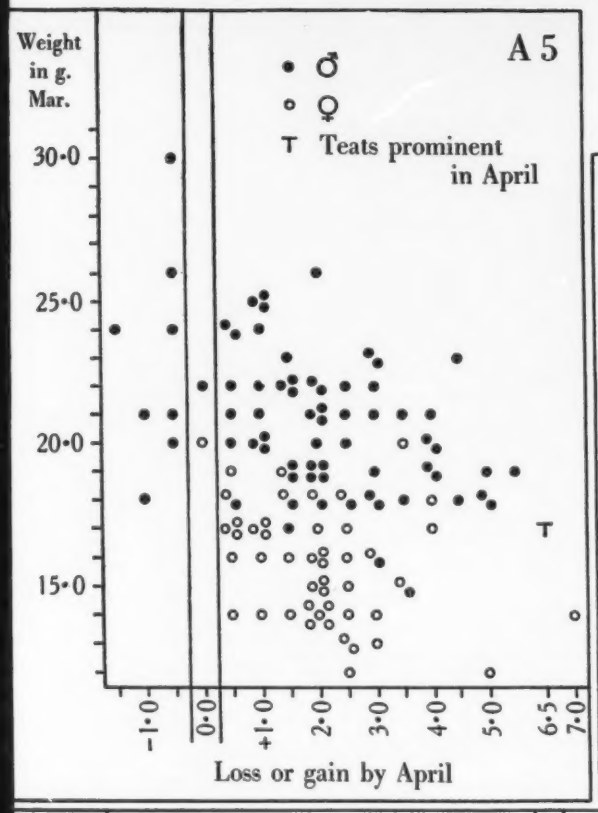

Weight
in g.
Nov.
D 5
25·0
20·0
15·0
10·0
10·0 15·0 20·0 25·0 30·0
Weight in April

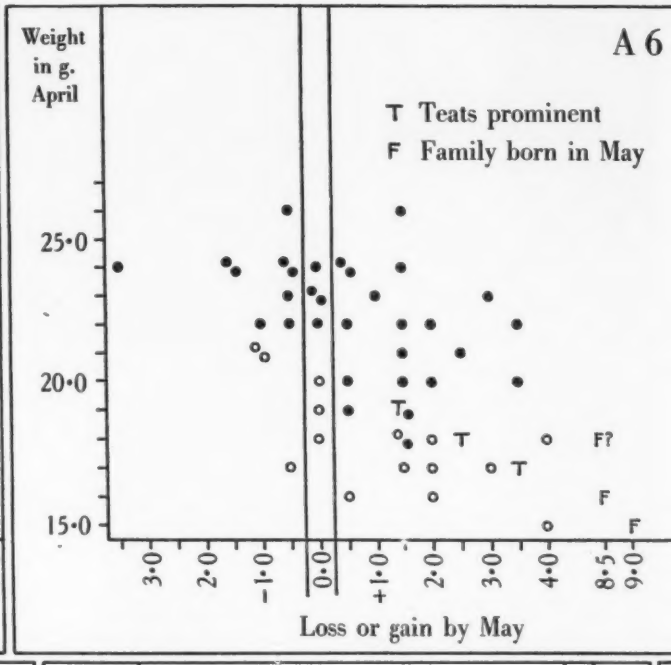

Weight
in g.
Jan.
20·0
15·0

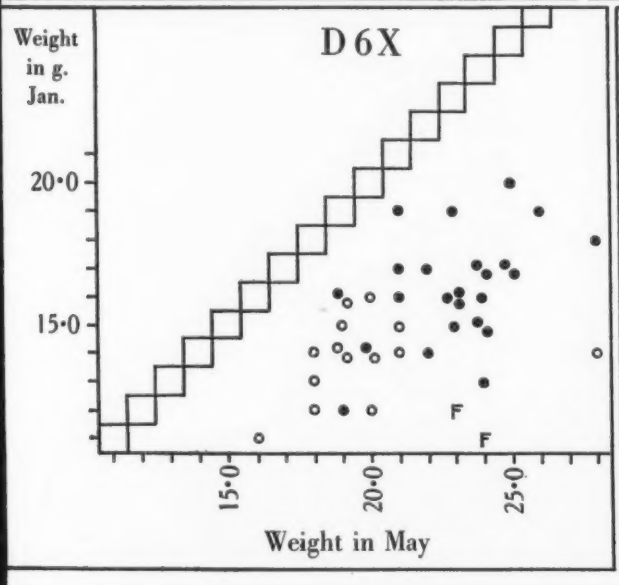

C 5X
T
6·0
7·0
pr.

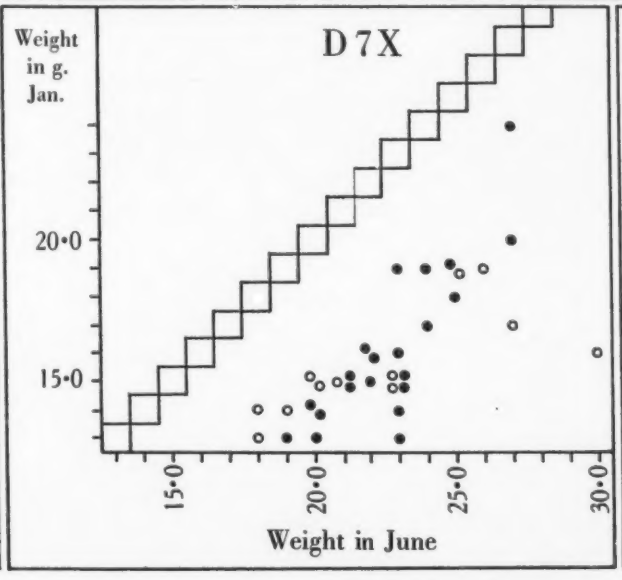

ight
g.
an.
D 5X
T
0·0
5·0
15·0
20·0
25·0
30·0
Weight in April

In June all ♀s had promin
marked T in diagrams.

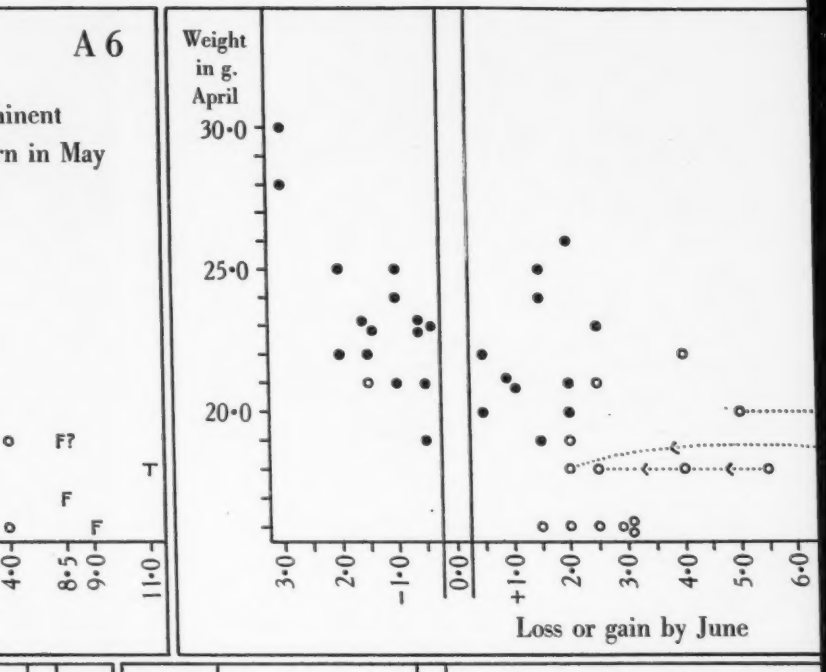


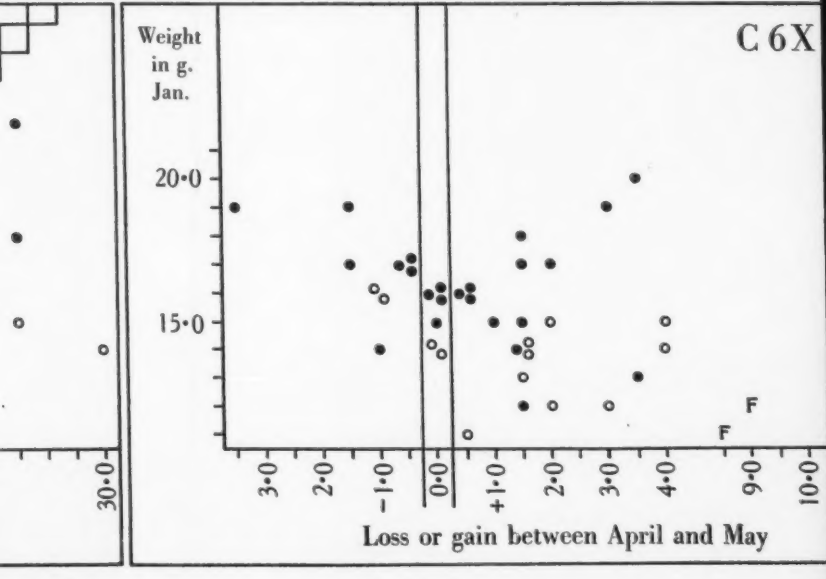






had prominent teats, although not
diagrams.





n June all ♀s had prominent teats, although not
narked **T** in diagrams.

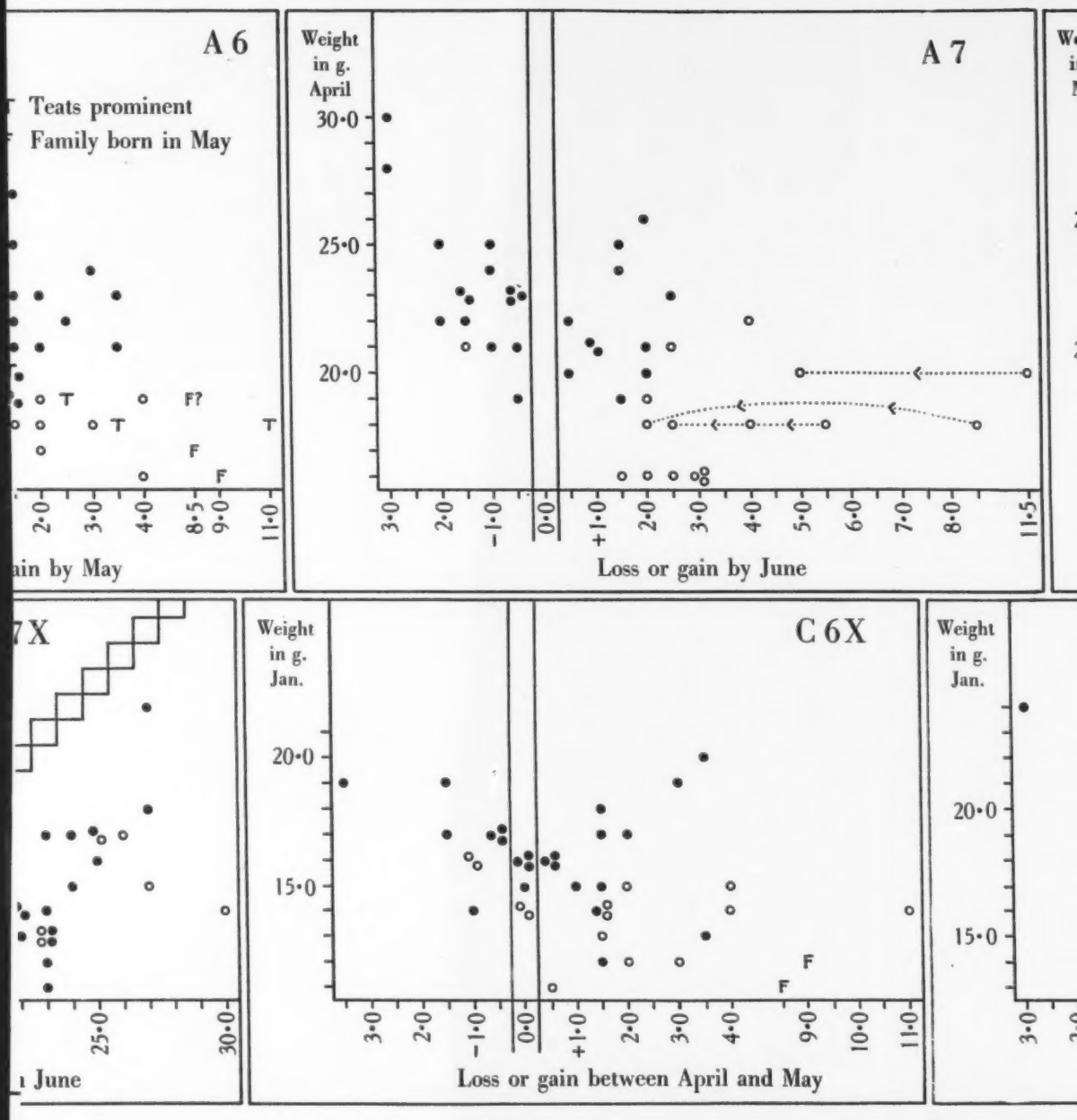
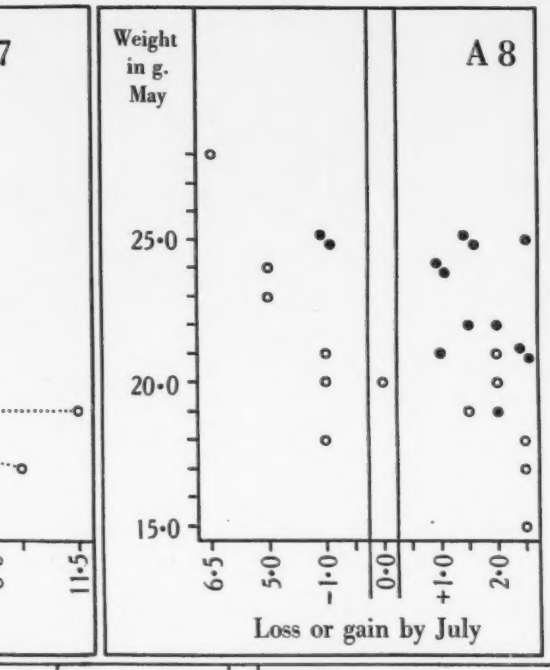
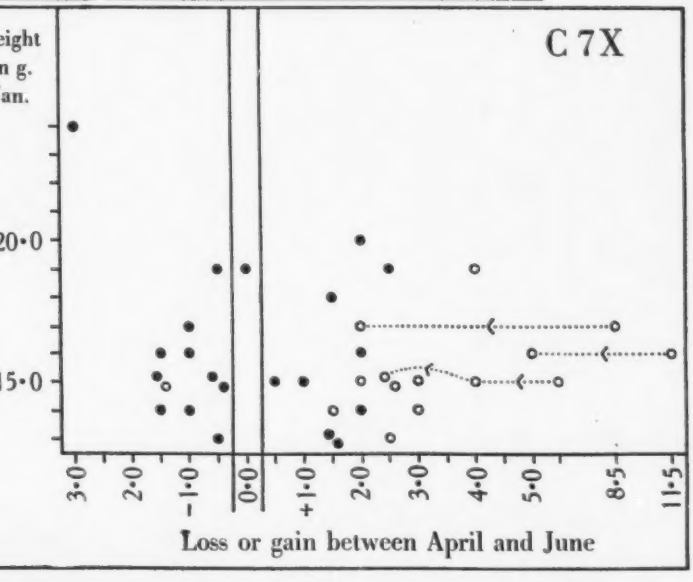


CHART 3


7
11·5

Weight in g. May
A 8
25·0
20·0
15·0
6·5
5·0
−1·0
0·0
+1·0
2·0
Loss or gain by July


eight
n g.
an.
C 7X
20·0
15·0
3·0
2·0
−1·0
0·0
+1·0
2·0
3·0
4·0
5·0
8·5
11·5
Loss or gain between April and June

THE CONTROL OF INDUSTRIAL PROCESSES SUBJECT TO TRENDS IN QUALITY

By L. H. C. TIPPETT, *British Cotton Industry Research Association*

In many industrial processes there is a progressive change in the quality of the products with time, which is tolerated up to a certain stage, after which corrective action is taken. An example occurs in the mass production of metal parts by a machining operation to some specified dimension. Random fluctuations in the dimension of articles produced at any one time are caused by such factors as variations in the materials processed and the manipulation of the machine, and superimposed on these is a trend due to tool wear, say in the direction of an increase in the dimension. At some stage, it is decided that too many articles are exceeding the tolerance limit of the dimension and something is done: we shall say that the tool is discarded.

In order to decide when to discard the tool, one procedure is to take samples at regular intervals of time, measure the articles, and calculate the mean dimension for the sample; when that mean reaches a certain value the tool is discarded. But the sample means are subject to sampling errors so that they do not measure exactly the 'true' or population mean of all the articles produced at the given time. Consequently, if tools are discarded at one level of sample means, some will be discarded before the population mean has reached that level and others afterwards; and there will be a frequency distribution of states of wear, as measured by the population mean, at which different tools are discarded. It is the purpose of this paper to determine that frequency distribution and discuss its practical consequences.

The argument will be developed in terms of the above example, but the results may be applied generally to any process having a trend in some characteristic that is estimated at regular intervals by any measure having a sampling error.

The situation is shown diagrammatically in Fig. 1. The line ABC represents the change with time in the population mean, X, of the dimension. X is a characteristic of the tool as set up in the machine, and we may imagine it to be determined, in principle, by making and measuring over a very short interval of time surrounding each instant a very large number of articles, and calculating their mean. In the region AB the trend may be of any form, but in the region BC, which covers the range of the population dimension at which substantially all the tools are discarded, the trend is assumed to be a straight line represented by the equation

$$X = d + ah, \tag{1}$$

where h is time measured in units of the intervals between the taking of successive samples, and d and a are constants. The constant d is chosen to be the level of sample means at which tools are discarded, so that h is measured from the time at which $X = d$. The instants at which samples are taken are

$$(\theta - s), \quad (\theta - s + 1), \quad (\theta - s + 2), \quad \ldots, \quad (\theta - 1), \quad \theta, \quad (\theta + 1), \quad \ldots,$$

where θ is the interval between zero time and the first subsequent sampling instant, and s is an integer defined below.

The sampling distribution of the mean is represented in Fig. 1 for the first three sampling instants, and

$$P(\theta-s), \quad P(\theta-s+1), \quad P(\theta-s+2)$$

are the respective probabilities that the sample means exceed d. For any sampling instant $(\theta-s+v)$ say, $P(\theta-s+v)$ is the probability that a tool, surviving to that instant, will then be discarded. Let $S(\theta-s+v)$ be the probability that any tool will survive to the instant $(\theta-s+v)$ and $D(\theta-s+v)$ the probability that it will be discarded at that instant, so that

$$D(\theta-s+v) = P(\theta-s+v)\, S(\theta-s+v). \tag{2}$$

We shall assume that for a population of tools, θ will have all values between 0 and 1 with equal probability and shall consider the elemental proportion $d\theta$ of those tools for which θ is within the limits $\theta \pm \frac{1}{2}d\theta$. We shall also choose s such that $P(\theta-s) = 0$ to a sufficient degree of accuracy (i.e. that $P(\theta-s) < \epsilon$, where ϵ is a chosen small quantity), and $P(\theta-s+1)$ is

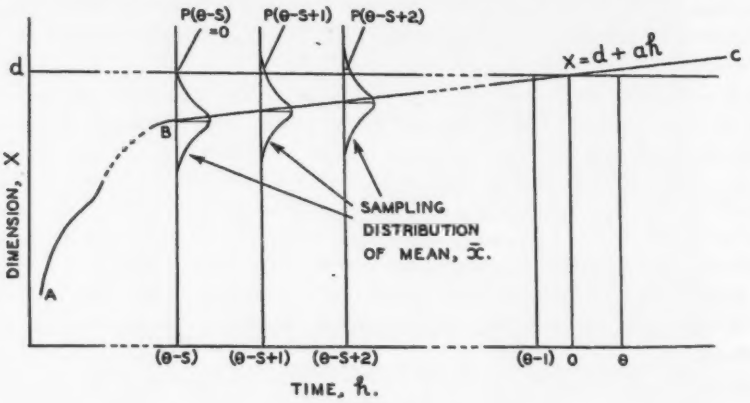


Fig. 1.

the smallest probability of which account need be taken. Then, with the aid of equation (2), we obtain for any tool in the elemental proportion the following probabilities of survival to and discard at the successive sampling instants:

Instant	Probability of survival to instant			Probability of discard at instant		
$(\theta-s)$	$S(\theta-s)$	$= 1$		$D(\theta-s)$	$= P(\theta-s)$	$S(\theta-s) \quad = 0$
$(\theta-s+1)$	$S(\theta-s+1)$	$= S(\theta-s)$	$-D(\theta-s)$	$D(\theta-s+1)$	$= P(\theta-s+1)$	$S(\theta-s+1)$
$(\theta-s+2)$	$S(\theta-s+2)$	$= S(\theta-s+1)$	$-D(\theta-s+1)$	$D(\theta-s+2)$	$= P(\theta-s+2)$	$S(\theta-s+2)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Thus, from the known probabilities $P(\theta-s)$, etc., can be deduced the probabilities $D(\theta-s)$, etc., and these multiplied by $d\theta$ are the elemental probabilities of the times of discard of tools, i.e. they are proportional to the ordinates of the probability distribution of the times of discard.

Probability distributions have been calculated for normal sampling distributions having a constant standard error (S.E.) given by the equation

$$\text{S.E.} = ka, \tag{3}$$

where $k = 1, 2, 5, 10, 20, 33{\cdot}\dot{3}$ and 50 respectively. The probabilities of the time of discard exceeding various values of h are given in Table 1; and a few useful constants, calculated

Table 1. *Probability of time of discard of tool being later than h*

$k=1$		$k=2$		$k=5$		$k=10$		$k=20$		$k=33\cdot3$		$k=50$	
h	Prob.	h	Prob.	h	Prob.	h	Prob.	h	Prob.	h	Prob.	h	Prob.
								−78	1·0000			−196·5	1·0000
						−38	1·0000	−76	0·9999	−129·5	1·0000	−192·5	0·9998
−3·4	1·0000	−6·8	1·0000	−18·5	1·0000	−37	0·9999	−74	0·9997	−126·5	0·9997	−188·5	0·9994
−3·2	0·9999	−6·4	0·9998	−18·0	0·9999	−36	0·9997	−72	0·9994	−123·5	0·9994	−184·5	0·9990
−3·0	0·9997	−6·0	0·9994	−17·5	0·9998	−35	0·9995	−70	0·9990	−120·5	0·9991	−180·5	0·9986
−2·8	0·9993	−5·6	0·9987	−17·0	0·9997	−34	0·9992	−68	0·9985	−117·5	0·9985	−176·5	0·9978
−2·6	0·9986	−5·2	0·9973	−16·5	0·9995	−33	0·9988	−66	0·9977	−114·5	0·9977	−172·5	0·9968
−2·4	0·9974	−4·8	0·9948	−16·0	0·9992	−32	0·9982	−64	0·9965	−111·5	0·9967	−168·5	0·9955
−2·2	0·9953	−4·4	0·9905	−15·5	0·9988	−31	0·9974	−62	0·9948	−108·5	0·9953	−164·5	0·9938
−2·0	0·9917	−4·0	0·9833	−15·0	0·9982	−30	0·9963	−60	0·9926	−105·5	0·9933	−160·5	0·9915
−1·8	0·9859	−3·6	0·9717	−14·5	0·9974	−29	0·9947	−58	0·9893	−102·5	0·9906	−156·5	0·9885
−1·6	0·9770	−3·2	0·9539	−14·0	0·9963	−28	0·9925	−56	0·9848	− 99·5	0·9870	−152·5	0·9845
−1·4	0·9637	−2·8	0·9279	−13·5	0·9948	−27	0·9895	−54	0·9788	− 96·5	0·9821	−148·5	0·9793
−1·2	0·9445	−2·4	0·8909	−13·0	0·9928	−26	0·9855	−52	0·9708	− 93·5	0·9756	−144·5	0·9725
−1·0	0·9178	−2·0	0·8407	−12·5	0·9901	−25	0·9802	−50	0·9603	− 90·5	0·9672	−140·5	0·9641
−0·8	0·8820	−1·6	0·7757	−12·0	0·9865	−24	0·9732	−48	0·9466	− 87·5	0·9561	−136·5	0·9532
−0·6	0·8359	−1·2	0·6954	−11·5	0·9819	−23	0·9640	−46	0·9288	− 84·5	0·9420	−132·5	0·9396
−0·4	0·7787	−0·8	0·6016	−11·0	0·9759	−22	0·9522	−44	0·9062	− 81·5	0·9240	−128·5	0·9226
−0·2	0·7105	−0·4	0·4984	−10·5	0·9681	−21	0·9372	−42	0·8777	− 78·5	0·9015	−124·5	0·9019
0·0	0·6325	0·0	0·3921	−10·0	0·9582	−20	0·9182	−40	0·8425	− 75·5	0·8735	−120·5	0·8764
0·2	0·5473	0·4	0·2904	− 9·5	0·9458	−19	0·8947	−38	0·7999	− 72·5	0·8392	−116·5	0·8457
0·4	0·4583	0·8	0·2006	− 9·0	0·9305	−18	0·8660	−36	0·7492	− 69·5	0·7981	−112·5	0·8090
0·6	0·3698	1·2	0·1279	− 8·5	0·9117	−17	0·8313	−34	0·6902	− 66·5	0·7495	−108·5	0·7660
0·8	0·2862	1·6	0·0745	− 8·0	0·8889	−16	0·7901	−32	0·6234	− 63·5	0·6934	−104·5	0·7163
1·0	0·2116	2·0	0·0392	− 7·5	0·8616	−15	0·7423	−30	0·5501	− 60·5	0·6299	−100·5	0·6600
1·2	0·1487	2·2	0·0273	− 7·0	0·8293	−14	0·6878	−28	0·4722	− 57·5	0·5601	− 96·5	0·5978
1·4	0·0988	2·4	0·0184	− 6·5	0·7918	−13	0·6270	−26	0·3923	− 54·5	0·4855	− 92·5	0·5304
1·6	0·0618	2·6	0·0120	− 6·0	0·7489	−12	0·5608	−24	0·3138	− 51·5	0·4085	− 88·5	0·4596
1·8	0·0362	2·8	0·0076	− 5·5	0·7006	−11	0·4907	−22	0·2401	− 48·5	0·3319	− 84·5	0·3873
2·0	0·0197	3·0	0·0047	− 5·0	0·6473	−10	0·4187	−20	0·1747	− 45·5	0·2589	− 80·5	0·3160
2·1	0·0141	3·2	0·0027	− 4·5	0·5896	− 9	0·3472	−18	0·1200	− 42·5	0·1928	− 76·5	0·2485
2·2	0·0099	3·4	0·0015	− 4·0	0·5284	− 8	0·2787	−16	0·0772	− 39·5	0·1362	− 72·5	0·1873
2·3	0·0068	3·6	0·0008	− 3·5	0·4650	− 7	0·2157	−14	0·0462	− 36·5	0·0905	− 68·5	0·1346
2·4	0·0046	3·8	0·0004	− 3·0	0·4010	− 6	0·1602	−12	0·0255	− 33·5	0·0562	− 64·5	0·0915
2·5	0·0030	4·0	0·0002	− 2·5	0·3381	− 5	0·1137	−10	0·0127	− 30·5	0·0322	− 60·5	0·0585
2·6	0·0019	4·2	0·0001	− 2·0	0·2780	− 4	0·0767	− 8	0·0057	− 27·5	0·0169	− 56·5	0·0349
2·7	0·0012	4·4	0·0000	− 1·5	0·2223	− 3	0·0490	− 6	0·0023	− 24·5	0·0080	− 52·5	0·0192
2·8	0·0007			− 1·0	0·1725	− 2	0·0295	− 4	0·0007	− 21·5	0·0034	− 48·5	0·0097
2·9	0·0004			− 0·5	0·1295	− 1	0·0166	− 2	0·0001	− 18·5	0·0012	− 44·5	0·0044
3·0	0·0002			0·0	0·0938	0	0·0087	0	0·0000	− 15·5	0·0003	− 40·5	0·0017
3·1	0·0001			0·5	0·0654	1	0·0043			− 12·5	0·0000	− 36·5	0·0006
3·2	0·0000			1·0	0·0437	2	0·0020					− 32·5	0·0001
				1·5	0·0279	3	0·0008					− 28·5	0·0000
				2·0	0·0170	4	0·0003						
				2·5	0·0098	5	0·0001						
				3·0	0·0054	6	0·0000						
				3·5	0·0028								
				4·0	0·0014								
				4·5	0·0007								
				5·0	0·0003								
				5·5	0·0001								
				6·0	0·0000								

from a fuller form of the distributions than that given in Table 1, are shown in Table 2. These constants involve the following quantities:

$\bar{h}$, the mean time of discard,

$h_{0\cdot01}$ and $h_{0\cdot05}$ the times of discard exceeded with probabilities of 0·01 and 0·05 respectively, and

$X_{0\cdot01}$ and $X_{0\cdot05}$ the population values of the dimension at the corresponding times, related to the values of h by equation (1).

Values from Table 2 are plotted against k in Figs. 1 and 2 to facilitate interpolation for intermediate values of k†.

Table 2

k	1	2	5	10	20	33·3	50
Mean time of discard, $\bar{h}$	0·27	−0·49	−4·01	−11·61	−29·68	−56·68	−93·16
Time exceeded with probability 0·01:							
$h_{0\cdot01}$*	2·20	2·68	2·50	−0·22	−9·38	−25·33	−48·66
$(h_{0\cdot01}-\bar{h})$	1·93	3·17	6·51	11·39	20·30	31·35	44·50
$(d-X_{0\cdot01})$/S.E.	−2·20	−1·34	−0·50	0·02	0·47	0·76	0·97
Time exceeded with probability 0·05:							
$h_{0\cdot05}$*	1·68	1·86	0·84	−3·04	−14·29	−32·82	−59·21
$(h_{0\cdot05}-\bar{h})$	1·41	2·35	4·85	8·57	15·39	23·86	33·95
$(d-X_{0\cdot05})$/S.E.	−1·68	−0·93	−0·17	0·30	0·71	0·98	1·18

* These are also $(X_{0\cdot01}-d)/a$ and $(X_{0\cdot05}-d)/a$ according to equation (1).

Applications

The results of the previous section will usually be applied to control the upper limit of X at which tools are discarded, for it is then that defective articles begin to be produced. The theory does not allow the specification of an absolute upper limit of X beyond which no tools will be used—the assumption of normality allows the theoretical possibility of tools being used up to the stage at which X approaches infinity. It is possible only to specify an upper limit of X that will be exceeded with some probability, and in the following discussion we shall take this probability to be 0·01 considering $X_{0\cdot01}$ to be the limit of X and $h_{0\cdot01}$ as the latest time that tools will be discarded.

First, we may calculate $X_{0\cdot01}$ for any chosen level of discard. The practice in ordinary quality control of setting control limits at 1·96 S.E. above and below the control level‡ suggests, on the face of it, that the level of discard, d, should be at sample means 1·96 S.E. below the upper tolerance limit L given in the specification of this article. Then

$$d = L - 1{\cdot}96\,\text{S.E.},$$

and if we subtract $d - X_{0\cdot01}$ from both sides we have

$$X_{0\cdot01} = L - \text{S.E.}\left\{\frac{d-X_{0\cdot01}}{\text{S.E.}} + 1{\cdot}96\right\}.$$

† The values in Tables 1 and 2 may have errors of 1 or, occasionally, 2 in the last figure.

‡ See Table 10 of pamphlet B.S. 600R, entitled, *Quality Control Charts*, by B. P. Dudding and W. J. Jennett, published by the British Standards Institution. The inner limits are referred to here.

From Table 2 we see that, if $k = 1$,

$$X_{0{\cdot}01} = L + 0{\cdot}24\,\text{S.E.}$$

Thus, more than one tool in a hundred are in use after they are producing articles with a population mean dimension above the tolerance limit, and hence with above 50 % defectives.

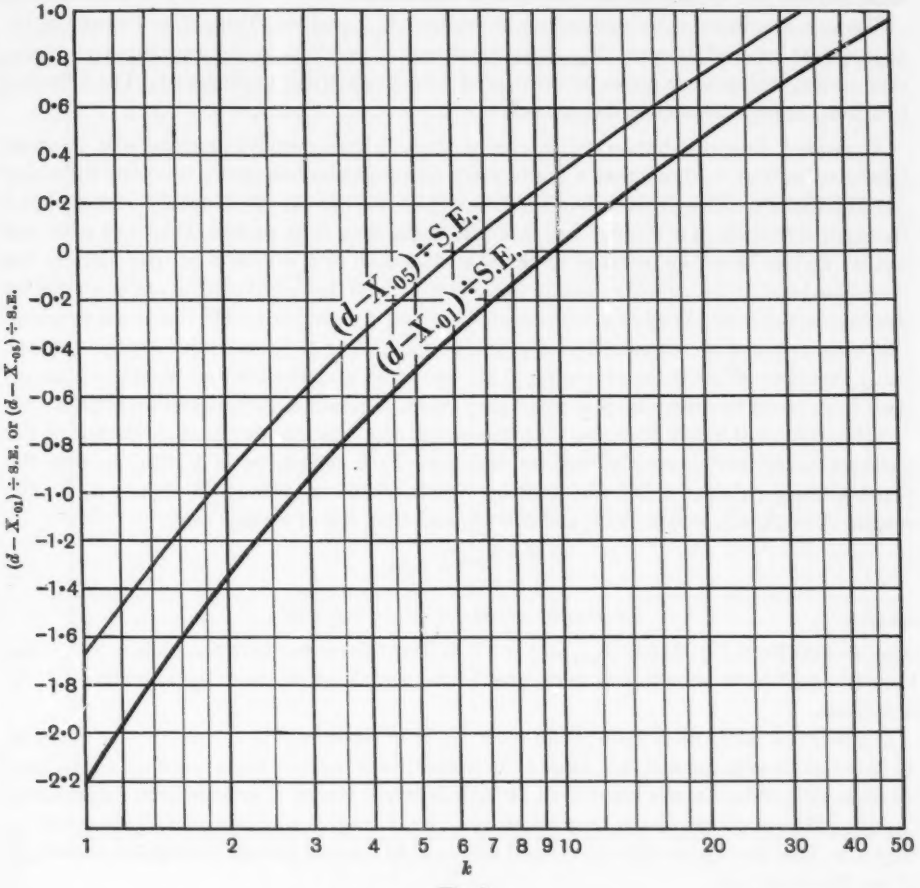


Fig. 2.

The situation is better if the rate of wear is greater compared with the standard error of sample means, so that k is increased. Thus, if $k = 50$,

$$X_{0{\cdot}01} = L - 2{\cdot}93\,\text{S.E.}$$

Now, in order to determine the percentage of defectives produced just before the time of discard, let us suppose that the sample size is 16 and that the standard deviation of individual articles made at any one time is σ, so that

$$\text{S.E.} = \tfrac{1}{4}\sigma.$$

Then, for $k = 50$,

$$X_{0{\cdot}01} = L - 0{\cdot}73\sigma,$$

and if the frequency distribution of individual articles made at any one time is normal, the percentage of defectives at the above value of $X_{0\cdot01}$ is about 22 (Table 9 of B.S. 600 R). This example is sufficient to show that a level of discard related to the tolerance limit in terms of only the standard error of the sample mean is unsatisfactory. The rate of change of X with time and the size of the sample must also be considered.

Usually, it will be more convenient to choose $X_{0\cdot01}$ and use Table 2 to determine the appropriate level of discard. $X_{0\cdot01}$ may be chosen in relation to the engineering limits so that at that value some given percentage of defective articles is produced. The following fictitious example illustrates the procedure.

Example 1. Suppose that an article is being produced to a specified diameter with an upper tolerance limit of 6·30 in., that a preliminary investigation has shown that the frequency distribution of articles produced on one machine at one time is substantially normal with a standard deviation of $\sigma = 0\cdot057$ in., that the variability is in control, that tool wear and similar factors cause an average increase in diameter of $a = 0\cdot0005$ in. per hour in the neighbourhood of the time of retooling and resetting, and that eight articles are measured for diameter every hour. At what sample mean diameter, d, must the tool be discarded to ensure that only one tool in a hundred ever produces more than 10 % defective articles through being oversize? What is the average tool life compared with the life that would be attained if all tools could be discarded just when they begin to produce 10 % defective articles?

Table 9 A of B.S. 600 R shows that the value of t, the standardized deviation of the variable distributed normally, corresponding to 10 % defectives is 1·2816, so that the corresponding value of X is $6\cdot30 - 1\cdot2816 \times 0\cdot057 = 6\cdot227$ in.; this is the chosen $X_{0\cdot01}$. For sample of eight, S.E. = 0·0201 in. and $k = 40$; and from Fig. 2 we find that

$$\frac{d - X_{0\cdot01}}{\text{S.E.}} = 0\cdot84,$$

whence

$$d = 0\cdot0201 \times 0\cdot84 + 6\cdot227 = 6\cdot244 \text{ in.}$$

Also we find from Fig. 3 that $(h_{0\cdot01} - \bar{h}) = 37$, so that the mean life of the tools is 37 hr. less than it would have been had it been possible to use them all until they produced 10 % defectives.

It should be noted that most of the tools (99 % of them) will never produce as many as 10 % defective articles, and that some (1 % of them) will produce more; for most of the time all tools will produce many fewer than 10 % defectives. Hence, if articles from a number of tools in different stages of wear can be bulked, the average percentage of defectives will be very low. This average can be determined if the (X, h) curve is known throughout the whole of the life of the tool.

The level of discard can be worked out for an $X_{0\cdot01}$ corresponding to any other limit of allowed percentage of defectives, or for a probability level other than 0·01 with which that limit is exceeded. The relevant quantities for a level of 0·05 are given in Table 2 and Figs. 1 and 2, and for other levels they may be obtained from Table 1. The choice of these quantities depends on technical conditions and requirements, and no guiding rules can be given. In practice it will be difficult to make a decision based on measured quantities, and a certain amount of guesswork and arbitrary choice will probably be involved. It would be easy, but laborious, to calculate the mean percentage of defective articles produced by all tools immediately before discard, for various levels of discard, and to use this in choosing the level; but it is doubtful if such a choice would be any easier to make in practice.

The above application considers the approach of articles to the upper tolerance limit only; the lower limit will determine the original setting of the tool in the machine. If the trend is one of decreasing dimension, the signs of all the quantities are reversed and the approach to the lower tolerance limit is controlled. The procedure does nothing towards controlling erratic fluctuations.

Example II. This example concerns the times to melt successive casts of steel in an open-hearth furnace, and the data are taken from *Statistical Methods in Industry*,* Fig. 7 and Table XX. In the figure, the melting times for successive casts are plotted against the cast

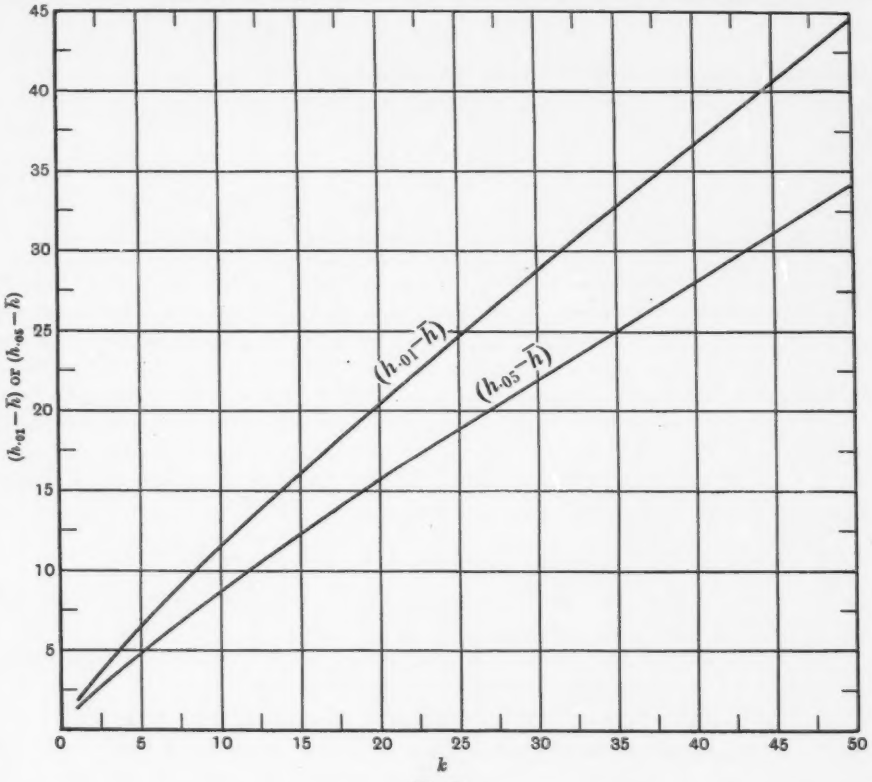


Fig. 3.

number, and for *A* furnace there is a substantial random fluctuation superimposed on a secular movement which appears to be roughly periodic, rising to peaks at intervals of 40–50 casts. A secular variation of this sort could be due to the gradual obstruction of the passages for the gases, which are cleared at intervals. The random variations could be due to such things as variations from cast to cast in the material charged into the furnace. It is not known whether these factors explain the fluctuations for *A* furnace, but for the sake of argument let us suppose that it is so, and let us derive a procedure for deciding when to clean the flues. The data are inadequate to justify this supposition and can provide only very

* Published by the Iron and Steel Industrial Research Council of the British Iron and Steel Federation.

rough estimates of the various quantities; but for the purposes of illustration this does not matter.

Ideally, we would describe the state of the furnace by the time it takes to melt an average cast, undisturbed by random fluctuations; this is our unknown population value, X. The standard deviation of the random fluctuations in the actual times about X is the S.E., and may be roughly estimated from the mean range between consecutive pairs of times. In this way, S.E. is estimated to be 1·17 hr. A rough estimate of a, the rate of increase in melting time, obtained from the figure is 0·16 hr. per cast, so that

$$k = \frac{1{\cdot}17}{0{\cdot}16} = 7$$

(h is the cast number), and from Fig. 2 we find that

$$(d - X_{0{\cdot}01})/\text{S.E.} = -0{\cdot}23,$$

whence

$$(d - X_{0{\cdot}01}) = -0{\cdot}27 \text{ hr.}$$

Suppose it is decided that $X_{0{\cdot}01}$ shall be 14 hr.; then $d = 13{\cdot}7$ hr. From Fig. 3 we find that $(h_{0{\cdot}01} - \bar{h}) = 7{\cdot}5$ casts, and the average melting time immediately before the furnace is cleaned is $14 - 7{\cdot}5 \times 0{\cdot}16 = 12{\cdot}8$ hr.

There are some conditions to be satisfied before the results of this paper can be applied. The assumption of the constancy of the standard error of sample means involves the assumption that the variability is in control. The mean dimension need not be in control, except for the trend, but it is better that it should be so. Otherwise the variability due to uncontrolled random variations in X must be added to the sampling variations when calculating the S.E. The question of control of X does not arise in circumstances like those envisaged in Example II, where the S.E. is due entirely to random variations in X. Another condition is that a and S.E. are known and that the sampling variation of $\bar{x}$, the sample means, is normal. All these conditions require a preliminary investigation.

Optimum conditions for control

It will be interesting to see whether it is better to take large samples at infrequent intervals or small ones at correspondingly more frequent intervals, the total number of tests over a given length of time remaining constant. For the probability level of 0·01, it is immaterial whether the intervals are frequent or infrequent provided that

$$(h_{0{\cdot}01} - \bar{h}) = \alpha k^{\frac{1}{2}},$$

where α is a constant. For if the interval between the samples is increased from 1 to r units of time and the sample size is increased to r times its original value, the new standard error is $\text{S.E.}' = \text{S.E.}/\sqrt{r}$, the rate of deterioration in terms of the new units of time is $a' = ra$, the new k becomes $k' = k/r^{\frac{1}{2}}$, and the new $(h_{0{\cdot}01} - \bar{h})$ becomes $\alpha k'^{\frac{1}{2}} = \alpha k^{\frac{1}{2}}/r$ in the new units, or $\alpha k^{\frac{1}{2}}$ in the original units of time, which is the same as the original value of $(h_{0{\cdot}01} - \bar{h})$. When the values of $(h_{0{\cdot}01} - \bar{h})$ in Table 2 are plotted against $k^{\frac{1}{2}}$, they are found to lie on a curve starting near the origin and concave to the $(h_{0{\cdot}01} - \bar{h})$ axis; i.e. $(h_{0{\cdot}01} - \bar{h})$ increases more rapidly than in proportion to $k^{\frac{1}{2}}$, and it is better to have a large value of r so as to keep k small. The same is true of $(h_{0{\cdot}05} - \bar{h})$. Hence, within the limits of this investigation, for a given total number of articles tested in a given total time, it is better, from the point of view of keeping the average tool life as near as possible to the chosen limiting life, to take large samples at relatively infrequent intervals than small samples at frequent intervals.

We may illustrate this by Example I. If samples of thirty-two articles are taken at 4-hourly intervals,

$$a = 4 \times 0{\cdot}0005 = 0{\cdot}0020 \text{ in. per interval},$$

$$\text{S.E.} = \frac{0{\cdot}057}{\sqrt{32}} = 0{\cdot}101 \text{ in.},$$

$$k = 5{\cdot}05,$$

and from Table II,

$$(h_{0{\cdot}01} - \bar{h}) = 6{\cdot}5 \text{ intervals}$$

$$= 26 \text{ hr. (approx.)}.$$

This may be compared with the 37 hr. for hourly samples of eight articles.

Another method of improving control that suggests itself is to combine each sample mean with the few previous means, using the assumption of a linear trend to obtain an improved estimate of X at the instant when the last sample is taken.

Let the number of means combined in this way be m, and let them be designated

$$\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_u, \ldots, \bar{x}_m,$$

$\bar{x}_m$ being the last in the series. Then it is easy to show that if a straight line is fitted to these values by the method of least squares, the improved estimate of X at the last sampling instant, $\bar{x}'_m$ say, is given by the equation

$$\bar{x}'_m = \frac{2}{m(m+1)} \sum_{u=1}^{u=m} (3u - m - 1)\bar{x}_u, \tag{4}$$

and the standard error of this estimate, S.E._m say, by the equation

$$\text{S.E.}_m = \sqrt{\frac{2(2m-1)}{m(m+1)}} \times \text{S.E.} \tag{5}$$

Table 3 shows how the precision of the estimate increases with m.

Table 3

m	$\text{S.E.}_m/\text{S.E.}$	m	$\text{S.E.}_m/\text{S.E.}$
1	1·00	5	0·77
2	1·00	6	0·72
3	0·91	8	0·65
4	0·84	10	0·59

It is wrong, of course, to combine so many means extending over such a length of time that the departure of the (X, h) curve from linearity for that time is appreciable compared with the sampling errors.

Suppose in Example I the hourly results are combined in this way in fours. Then

$$\text{S.E.}_m = 0{\cdot}84 \times 0{\cdot}0201 = 0{\cdot}0169 \text{ in.},$$

$$k = 33{\cdot}8,$$

and

$$(h_{0{\cdot}01} - \bar{h}) = 31 \text{ hr. (approx.)}.$$

For this example, the concentration of the thirty-two tests at 4-hourly intervals, giving $(h_{0{\cdot}01} - \bar{h}) = 26$ hr., remains the best arrangement. Sometimes, however, it is better to test more frequently and combine means by this second method. For example, if with hourly

tests, $k = 31{\cdot}62$, then $(h_{0{\cdot}01} - \bar{h}) = 30$ hr. If ten times as many tests are taken at 10-hourly intervals, k becomes $1{\cdot}0$ and $(h_{0{\cdot}01} - \bar{h}) = 10 \times 2{\cdot}20 = 22$ hr. If the hourly tests are combined in tens, k becomes $0{\cdot}59 \times 31{\cdot}62 = 18{\cdot}6$, and $(h_{0{\cdot}01} - \bar{h}) = 19$ hr. This second method of combination is the only one available in cases like that of Example II.

Summary

1. If samples of a product deteriorating linearly with time are taken at intervals, and action is taken when the sample values reach a certain level, such action will be taken, sometimes before and sometimes after the population value has reached that level.

2. Frequency distributions are given for the times at which action is taken. These distributions depend on the standard error of the sample mean, the size of the sample and the rate of deterioration of the product.

3. A procedure is given for deciding when to take action.

4. For a given number of tests, it is better to take large samples infrequently than small ones frequently. A method is given for combining the results of several consecutive samples.

STUDENTIZATION

OR THE ELIMINATION OF THE STANDARD DEVIATION OF THE PARENT POPULATION FROM THE RANDOM SAMPLE-DISTRIBUTION OF STATISTICS

By H. O. HARTLEY

1. Introduction

In this paper we give a mathematical solution of a problem in statistical distribution, sometimes referred to as 'Studentization'. This process is best described by recalling briefly the properties of two well known examples:

Let $x_1, \ldots, x_n$ denote a sample of n observations drawn from a normal distribution with mean μ and standard deviation σ. Then the difference between sample and population mean, $\bar{x}-\mu$, is normally distributed with zero mean and standard deviation $\sigma/\sqrt{n}$. If we estimate the standard deviation of the parent, σ, by

$$s = \sqrt{[\Sigma(x-\bar{x})^2/(n-1)]},$$

then the ratio

$$t = (\bar{x}-\mu)\sqrt{n}/s$$

is distributed according to Student's t-distribution which does not depend on σ.

Consider now a second, independent sample $X_1, \ldots, X_m$ drawn from the same population. The random sample distribution of its standard deviation

$$s^* = \sqrt{[\Sigma(X-\bar{X})^2/(m-1)]}$$

follows a well-known probability law which, again, involves the value of σ. If, again, we estimate this value of σ by using the standard deviation, s, of the first sample as its estimate, then the distribution of the ratio

$$\sqrt{F} = s^*/s$$

is independent of σ. This distribution is, of course, a transformation of Fisher's z-distribution.

It will be noted that the original statistics, $\bar{x}-\mu$ and s^*, as well as s, are proportional to σ in the sense that if the observations x_i and X_j are measured in units of σ the above statistics are, by this change of scale, transformed into $(\bar{x}-\mu)/\sigma$, s^*/σ and s/σ respectively, and their parental distribution is transformed into the normal distribution with unit standard deviation. Hence the random sample distributions of the ratios

$$t = \frac{\bar{x}-\mu}{\sigma} : \frac{s}{\sqrt{n}\,\sigma} = \frac{(\bar{x}-\mu)}{s}\sqrt{n}, \quad \sqrt{F} = \frac{s^*}{\sigma} : \frac{s}{\sigma} = \frac{s^*}{s}$$

do not involve the value of σ, nor is it necessary to know this value of σ in order to calculate the above ratios. The latter may be called the studentized statistics corresponding to the original statistics $\bar{x}-\mu$ and s^*.

This simple studentization process of dividing a statistic by an independent estimate s of the parental standard deviation σ may be applied to a number of other useful statistics which are also proportional to σ in the above sense. We may mention here the range in a sample, the extreme observation, the extreme value in a sample of random standard deviations, but there are many others.

It is the object of this paper to derive a general formula for the studentized distribution law of such statistics. The accuracy of the formula will be demonstrated with a numerical

example and a new formula for the incomplete Beta function derived in the process. This result has already been used in recent numerical work on this function (Thompson 1941). A second application of the method may be found in a paper by Pearson & Hartley (1943), and it is hoped to publish others.

Finally, computational methods for the systematic tabulation of studentized distribution laws are briefly outlined.

2. Definitions and notation

A problem akin to the present one has been considered in a previous paper (Hartley, 1938), but to fix ideas we shall repeat here certain definitions and results given there.

Let $x_1, \ldots, x_n$ be a sample of n observations drawn from a normal population with standard deviation σ. Consider a general statistic W, calculated from this sample, having the following properties:

(a) W is positive.†

(b) W is 'proportional to σ' in the sense that if the observations x_i are measured in units of σ the statistic W is thereby transformed into W/σ.

Let $f(W)$ be the distribution function of W for the case where σ is known to be unity, and let $P(W)$ be the corresponding probability integral, i.e. $P(W) = \int_0^W f(W)\,dW$. Also let s be an independent estimate of the standard deviation based on n degrees of freedom. Finally, let‡

$$S = \sqrt{n}\,s = \sqrt{(\text{Sum of squares})} \quad \text{and} \quad r = W/S.$$

Further denote by $f_n(r)$ the random sample distribution of r and by $p_n(R)$ its probability integral, i.e.

$$p_n(R) = \int_0^R f_n(r)\,dr.$$

Using this notation we proved (see Hartley, 1938) that

$$p_n(R) = \Gamma(\tfrac{1}{2}n)^{-1}\,2^{-\frac{1}{2}n+1}\int_0^\infty S^{n-1}e^{-\frac{1}{2}S^2}\,P(SR)\,dS. \tag{1}$$

This equation gives $p_n(R)$, the 'studentized integral', formally in terms of the '∞-integral' $P(W)$. We also proved the recurrence formula

$$p_{n-2}(R) = (n-2)\,R^{-(n-2)}\int_0^R p_n(r)\,r^{n-3}\,dr, \tag{2}$$

with the help of which the degrees of freedom of $p_n(R)$ may be reduced by 2.

3. Properties of the recurrence formula

The series of studentized integrals $p_n(R)$ is uniquely determined by the recurrence formula (2) and the relations

$$\lim_{n\to\infty} p_n\left(\frac{W}{\sqrt{n}}\right) = P(W) \quad \text{(uniformly in } W\text{)}, \quad p_n \text{ bounded}, \tag{3}$$

which follow from an inequality derived previously (see Hartley, 1938, (10)) and from (2). To prove this suppose there were two series of piecewise continuous functions $p_n(R)$ and

† It will be seen that this condition is not essential for our results. Also, if the distribution function of W is symmetrical about 0, condition (a) will be satisfied for the statistic $|W|$ which may be used in place of W.

‡ It was necessary to alter the notation used previously (Hartley, 1938) writing S in place of ρ and r in place of q, in order to fall in with the notation used by other writers (e.g. Newman, 1939) who reserve q for the ratio W/s.

$\pi_n(R)$ both satisfying (2) and (3), and let us denote by Δ_n the maximum of $|p_n(R)-\pi_n(R)|$ for $0 \leqslant R < \infty$. From (3) we obtain

$$\lim_{n\to\infty} \Delta_n = 0. \tag{4}$$

On the other hand, we find from equation (2) that

$$p_{n-2}(R)-\pi_{n-2}(R) = (n-2)\,R^{-(n-2)}\int_0^R [p_n(r)-\pi_n(r)]\,r^{n-3}\,dr,$$

from which we obtain $\Delta_{n-2} \leqslant (n-2)\,R^{-(n-2)}\Delta_n \dfrac{R^{n-2}}{n-2} = \Delta_n,$

i.e.

$$\Delta_{n-2} \leqslant \Delta_n. \tag{5}$$

From (4) and (5) it follows that

$$\Delta_n = 0$$

for all n.

The recurrence formula (2) may be used to produce by quadrature $p_{n-2}(R)$ from $p_n(R)$. The difficulty is that none of the integrals $p_n(R)$ is known so that we have no starting point. All we know is that

$$\lim_{n\to\infty} p_n\left(\frac{W}{\sqrt{n}}\right) = P(W).$$

This suggests that a first and main step of the process will consist in bridging the gap between $n = \infty$ (i.e. $P(W)$) and moderate values of n (i.e. $p_n(R)$), say, covering the range between 10 and 50. This step will be accomplished with the help of the theory of partial differential equations in the next section.

4. The partial differential equation of $p_n(R)$

We may write formula (2) in the form

$$R^n p_n(R) = n\int_0^R p_{n+2}(r)\,r^{n-1}\,dr. \tag{6}$$

Differentiating with regard to R we obtain

$$R\frac{dp_n}{dR} - n[p_{n+2}(R)-p_n(R)] = 0. \tag{7}$$

We now regard the degrees of freedom n as a continuous second variable capable of attaining positive values and try to find a function of two variables $p(n, R)$ for which the relations

$$R\frac{\partial p(n,R)}{\partial R} - n[p(n+2,R)-p(n,R)] = 0 \tag{8}$$

and

$$\lim_{n\to\infty} p\left(n, \frac{W}{\sqrt{n}}\right) = P(W) \tag{9}$$

hold. Again this function $p(n, R)$ is uniquely determined by the relations (8) and (9). This is easily seen from integrating (8) which yields

$$R^n p(n,R) = n\int_0^R p(n+2,r)\,r^{n-1}\,dr + g(n).$$

Putting $R = 0$ we see that $g(n) \equiv 0$, i.e. that equation (6) holds. The remainder follows by the argument given in § 3.

To convert equation (8) into a partial differential equation we expand the finite difference in a Taylor series and write

$$R\frac{\partial p}{\partial R}-n\left[2\frac{\partial p}{\partial n}+2\frac{\partial^2 p}{\partial n^2}+\frac{4}{3}\frac{\partial^3 p}{\partial n^3}+\frac{2}{3}\frac{\partial^4 \overline{p}}{\partial n^4}\right]=0, \tag{10}$$

where p is taken at arguments R and n whilst $\overline{p}$ is taken at R and some mean value between n and $n+2$.

We now introduce as new variables

$$y=\log R \quad \text{and} \quad x=\log n.$$

The partial differential equation is thereby transformed into

$$\frac{\partial p}{\partial y}-2\frac{\partial p}{\partial x}=2e^{-x}\left(\frac{\partial^2 p}{\partial x^2}-\frac{\partial p}{\partial x}\right)+\tfrac{4}{3}e^{-2x}\left(\frac{\partial^3 p}{\partial x^3}-3\frac{\partial^2 p}{\partial x^2}+2\frac{\partial p}{\partial x}\right)$$
$$+\tfrac{2}{3}e^{-3x}\left(\frac{\partial^4 \overline{p}}{\partial x^4}-6\frac{\partial^3 \overline{p}}{\partial x^3}+11\frac{\partial^2 \overline{p}}{\partial x^2}-6\frac{\partial \overline{p}}{\partial x}\right), \tag{11}$$

where the arguments of p are y and x and those of $\overline{p}$ are y and some mean value x between x and $\log(e^x+2)$.

5. The solution of the partial differential equation by iteration

The partial differential equation (11) cannot, of course, be solved by standard methods. However, it is possible to find an approximate solution for large and moderate values of x with the help of an iteration process. For $x\to\infty$ the equation (11) tends to the limiting form

$$\frac{\partial p}{\partial y}-2\frac{\partial p}{\partial x}=0, \tag{12}$$

the general solution of which is given by

$$p_0(x,y)=\phi(y+\tfrac{1}{2}x),$$

where $\phi(\lambda)$ is an arbitrary differentiable function of λ. In order to satisfy condition (9) we have to put

$$\phi(\lambda)=P(e^{\lambda}), \tag{13}$$

so that we obtain for the solution $p_0(x,y)$

$$p_0(x,y)=P(e^{y+\frac{1}{2}x}). \tag{14}$$

The first approximation $p_0(x,y)$ to the studentized integral† is therefore a function of $(y+\frac{1}{2}x)$ only i.e., expressed in terms of R and n, p_0 depends on $R\sqrt{n}$ only. This means that for large n the studentized integral is, to a first approximation, independent of n.

In order to work out closer approximations we have to solve the non-homogeneous partial differential equation

$$\frac{\partial p}{\partial y}-2\frac{\partial p}{\partial x}=\psi(y,x). \tag{15}$$

A partial solution of this equation can be found with the help of the theory of characteristics and is given by

$$p(x,y)=\int_0^y \psi(\rho,-2\rho+2y+x)\,d\rho, \tag{16}$$

† We shall confine ourselves here to working out the first three steps of the iteration process. It would lead us too far afield if we would give here the proof for its convergence. The check on the accuracy of the approximation is discussed in the last section.

as is easily verified by differentiation. The general integral of equation (15) is therefore given by

$$p(x,y)=\int_{y_0}^{y}\psi(\rho,-2\rho+2y+x)\,d\rho+\phi(y+\tfrac{1}{2}x), \tag{17}$$

where ϕ, y_0 are arbitrary and ρ is a variable of integration. If the function $\psi(y,x)$ is of the special form

$$\psi(y,x)=\sum_{\nu=1}^{N}e^{-\nu x}\psi_\nu(y+\tfrac{1}{2}x), \tag{18}$$

and if we put $y_0=-\infty$ we find by substituting (18) in (17) and by integrating each term

$$p(x,y)=\phi(y+\tfrac{1}{2}x)+\sum_{\nu=1}^{N}\frac{e^{-\nu x}}{2\nu}\psi_\nu(\tfrac{1}{2}x+y). \tag{19}$$

The second approximation ($p_1(x,y)$) to the solution of (11) is now obtained by solving

$$\frac{\partial p_1}{\partial y}-2\frac{\partial p_1}{\partial x}=2e^{-x}\left(\frac{\partial^2 p_0}{\partial x^2}-\frac{\partial p_0}{\partial x}\right), \tag{20}$$

using $p_0(x,y)=\phi(y+\frac{1}{2}x)$. From (15), (18) and (19) we obtain

$$p_1(x,y)=\phi(y+\tfrac{1}{2}x)+e^{-x}(\tfrac{1}{4}\phi''-\tfrac{1}{2}\phi'), \tag{21}$$

where we denote by $'$, $''$, $'''$, etc., the order of the derivative of the function ϕ with regard to its argument. The function $p_1(x,y)$ is a solution of the equation (11) if all terms involving e^{-2x}, e^{-3x}, etc., are ignored.

The third approximation $p_2(x,y)$ is now obtained as a solution of the equation

$$\frac{\partial p_2}{\partial y}-2\frac{\partial p_2}{\partial x}=2e^{-x}\left(\frac{\partial^2 p_1}{\partial x^2}-\frac{\partial p_1}{\partial x}\right)+\tfrac{4}{3}e^{-2x}\left(\frac{\partial^3 p_1}{\partial x^3}-3\frac{\partial^2 p_1}{\partial x^2}+2\frac{\partial p_1}{\partial x}\right), \tag{22}$$

where, in the right-hand side, we may ignore all terms involving e^{-3x}, e^{-4x}, etc. From equations (18) and (22) we obtain for the solution (19)

$$p_2(x,y)=\phi+e^{-x}(\tfrac{1}{4}\phi''-\tfrac{1}{2}\phi')+e^{-2x}(\tfrac{1}{32}\phi^{\mathrm{IV}}-\tfrac{5}{24}\phi'''+\tfrac{3}{8}\phi''-\tfrac{1}{6}\phi'), \tag{23}$$

which is a solution of (11) if all terms involving e^{-3x}, e^{-4x}, etc. (i.e. $1/n^3$, $1/n^4$, etc.), are ignored. For samples of the order 20 or larger this formula will in most cases give quite sufficient accuracy. For smaller samples or high accuracy the term involving e^{-3x} will be required. We find by a further iteration

$$p_3(x,y)=p_2(x,y)+\frac{e^{-3x}}{12}(\tfrac{1}{32}\phi^{\mathrm{VI}}-\tfrac{21}{48}\phi^{\mathrm{V}}+\tfrac{17}{8}\phi^{\mathrm{IV}}-\tfrac{11}{6}\phi'''+3\phi''). \tag{24}$$

Let us write formula (23) in terms of the statistic

$$R\sqrt{n}=e^{(y+\frac{1}{2}x)}.$$

We have the following relations between the function $\phi(\lambda)$ and the ∞-integral $P(W)$ ($\lambda=y+\frac{1}{2}x$)

$$\begin{aligned}
\phi &= P(e^{\lambda}),\\
\phi' &= P'e^{\lambda},\\
\phi'' &= P''e^{2\lambda}+P'e^{\lambda},\\
\phi''' &= P'''e^{3\lambda}+3P''e^{2\lambda}+P'e^{\lambda},\\
\phi^{\mathrm{IV}} &= P^{\mathrm{IV}}e^{4\lambda}+6P'''e^{3\lambda}+7P''e^{2\lambda}+P'e^{\lambda}.
\end{aligned}$$

Substituting in equation (23) we obtain the chance $p_n(R)$ for the statistic $r = W/s\sqrt{n}$ to fall below R which is identical with the chance for the studentized statistic W/s to fall below $R\sqrt{n} = Q$ (say). We obtain from the second approximation, i.e. from formula (23)

$$p_n(R) = P(Q) + \frac{1}{4n}(Q^2P'' - QP') + \frac{1}{16n^2}(\tfrac{1}{2}Q^4P^{\text{iv}} - \tfrac{1}{3}Q^3P''' - \tfrac{1}{2}Q^2P'' + \tfrac{1}{2}QP'), \qquad (25)$$

where the argument of P and its derivatives is Q, since

$$\phi(y + \tfrac{1}{2}x) = P(e^{y+\frac{1}{2}x}) = P(R\sqrt{n}) = P(Q).$$

With equations (25) and (24) we have reached the solution of the studentization which is of sufficient accuracy for most practical purposes. In the following sections we shall show with an example how these equations may be used in practical instances. We shall see that the formulae lend themselves very well to the computation of $p_n(R)$. For a fixed value of Q $p_n(Q/\sqrt{n})$ is a quadratic in $1/n$ if equation (25) is used. The formula is of a fair accuracy even for small values of n. However, if higher accuracy is required it is preferable to include the term involving $1/n^3$ which is given by equation (24).

6. A new formula for the incomplete B-function†

As is well known the incomplete B-function, the probability integral of Fisher's z and of Snedecor's F, are all transformations of the same probability law which is the studentized integral of a sample standard deviation s_1. If s_1^2 and s_2^2 are both estimates of σ^2 based on n_1 and n_2 degrees of freedom respectively, then the probability that the ratio s_1^2/s_2^2 falls below F is given by

$$I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2),$$

where I_{x_0} is the incomplete B-function and

$$x_0 = \frac{n_1 F}{n_2 + n_1 F}. \qquad (26)$$

Whilst for large p and q there are a number of useful methods facilitating the approximate evaluation of $I_x(p, q)$ (see e.g. Soper, 1921 and Wishart, 1927) there appears to be a lack of suitable methods if p is small or moderate and q is large. This is a case where our method of studentization may be usefully employed.

The distribution of s_1 for n_1 degrees of freedom is given by

$$f(s_1) = \{\Gamma(\tfrac{1}{2}n_1)\}^{-1} 2^{-\frac{1}{2}n_1+1} n_1^{n_1} s_1^{n_1-1} e^{-\frac{1}{2}s_1^2 n_1}. \qquad (27)$$

All we have to do therefore is to substitute in equation (25) $f(s_1)$ for P' and the derivatives of $f(s_1)$ for P'', P''' and P^{iv} respectively. The argument Q is to be replaced by the square root of the variance ratio F and n by n_2, whilst P itself is the probability integral of $\sqrt{(\chi^2/n_1)}$ for n_1 degrees of freedom.

We find without difficulty

$$p_{n_2}(\sqrt{(Fn_2)}) = P^*(\sqrt{F}) + \frac{1}{4n_2}[Ff' - \sqrt{F}f] + \frac{1}{16n_2^2}(\tfrac{1}{2}F^2f''' - \tfrac{1}{3}F^{\frac{1}{2}}f'' - \tfrac{1}{2}Ff' + \tfrac{1}{2}\sqrt{F}f), \qquad (28)$$

where f is given by (27) and the functions f', f'' and f''' are its derivatives whilst the argument s_1 is replaced by $\sqrt{F}$. The integral P^* is the probability integral of s_1 or that of $\sqrt{(\chi^2/n_1)}$.

† Since this paper was written my attention has been drawn to a paper by A. G. Campbell (1923) in which a similar expression has been derived. Campbell's formula is mentioned and used in the recent book by Simon (1941).

It is most convenient to adopt the notation which has been introduced by Karl Pearson in his work on the incomplete B-function. The integral p_n will then become an approximation to the function $I_{x_0}(p, q)$ where $2p = n_1$, $2q = n_2$ and x_0 is given by (26). If, further,

$$\omega = \tfrac{1}{2}Fn_1 = \frac{qx_0}{1-x_0}, \tag{29}$$

we obtain from equations (27) and (28) the new formula for the function $I_{x_0}(p, q)$, viz.

$$I_{x_0}(p,q) \equiv P^*(\sqrt{F}) + \tfrac{1}{2}e^{-\omega}\omega^p q^{-1}\Gamma(p)^{-1}(p-1-\omega) + \tfrac{1}{4}e^{-\omega}\omega^p q^{-2}\Gamma(p)^{-1}\{\tfrac{1}{2}p^3 - \tfrac{5}{2}p^2 + \tfrac{3}{2}p - \tfrac{1}{3} - \omega(\tfrac{3}{2}p^2 - \tfrac{11}{6}p + \tfrac{1}{3}) + \omega^2(\tfrac{3}{2}p - \tfrac{1}{6}) - \tfrac{1}{2}\omega^3\}, \tag{30}$$

where $P^*(\sqrt{F})$ is the probability integral of $\sqrt{(\chi^2/n_1)}$ for $2p$ degrees of freedom.

The formula is of considerable help for large values of q and small or moderate values of p. If p and ω are kept constant $I_{x_0}(p, q)$ is seen to be a quadratic in $1/q$.

To give some idea of the remarkable accuracy of the formula we consider the example $p = 1, \omega = 5{\cdot}555$ and $q = 5(1)\,10(2{\cdot}5)\,25(5)\,50$. A comparison of accurate value and approximation is given in the table below. The method has been used to calculate values of the incomplete B-function for $q = 120$ and small values of p. These were required for the new table of percentage levels of $I_{x_0}(p, q)$ prepared by *Biometrika* (Thompson, 1941).

$2q$	$I_x(1, q)$		$2q$	$I_x(1, q)$	
	Exact	Approx.		Exact	Approx.
100	0·994 8462	0·994 8452	30	0·991 140	0·991 093
90	6916	6901	25	0·989 922	0·989 830
80	4952	4930	20	0·987 95	0·987 77
70	2374	2342	18	0·986 79	0·986 55
60	0·993 8847	0·993 8795	16	0·985 29	0·984 94
50	374	365	14	0·983 26	0·982 73
45	022	010	12	0·980 41	0·979 54
40	0·992 572	0·992 553	10	0·976 26	0·974 63
35	0·991 973	0·991 943			

7. SOME REMARKS ON THE TABULATION OF 'STUDENTIZED INTEGRALS'

Formulae (24) and (25) are really the first three or four terms of an expansion of the studentized integral in powers of $1/n$, and it is necessary to know something about the accuracy of this approximation. The neglected remainder of the expansion depends on the derivatives of the ∞-integral $P(W)$ and it is difficult to reach a general formula which may be used as a convenient gauge for the estimation of its magnitude. Nevertheless, it will be possible to control, numerically, the accuracy of formula (25). Thus, if it is desired to tabulate the studentized integral to a certain decimal accuracy (say 3 or 4 or 5 decimal accuracy) one would proceed on the following lines:

(i) From the magnitudes of the coefficients of n^0, n^{-1}, n^{-2} and n^{-3} in equations (25) and (24) *estimate roughly* the smallest value of n (n' say) for which equation (25) represents $p_n(R)$ sufficiently accurately for large values of R.

(ii) For two or three values of n in the neighbourhood of the above value n' and paired with two or three values of R, calculate the exact studentized integral $p_n(R)$ from equation (1) by numerical quadrature and compare the result with the value of $p_n(R)$ given

by (25). This comparison should *fix* the smallest value of n (n_0 say) for which the approximate formula (25) yields the desired decimal accuracy.

(iii) For $n \geqslant n_0+2$ use formula (25) for the tabulation of $p_n(R)$. For $n < n_0+2$ use the recurrence formula (2) (starting from n_0+2 and n_0+1 respectively) to produce in turn the exact $p_n(R)$ by numerical quadrature† reducing the degrees of freedom by 2 in each step of the recurrence. The first step of this process should produce the studentized integral $p_n(R)$ and this should agree identically with the one calculated from formula (25).

Finally, we should mention a simple transformation of the recurrence formula (2) which makes it possible to carry out the quadrature (iii) with the help of a mechanical integrator known as the 'Differential Analyser'. A good description of this machine is given by Hartree (1935). We first transform the R-scale by introducing $\rho = R^{-2}$. We then replace the function $p_n(R) = p_n(\rho^{-\frac{1}{2}})$ by

$$\pi_n(\rho) = [1 - p_n(\rho^{-\frac{1}{2}})]\rho^{-\frac{1}{2}n}. \tag{31}$$

The recurrence formula (2) is thereby transformed into

$$\pi_n(\mathrm{P}) = \tfrac{1}{2}n\int_{\mathrm{P}}^{\infty} \pi_{n+2}(\rho)\,d\rho. \tag{32}$$

Thus $\pi_n(\rho)$ is seen to be a multiple of the integral of $\pi_{n+2}(\rho)$ and the recurrence process is reduced to ordinary integration.

This process is easily carried out on the Differential Analyser. For instance, with a capacity of six 'integrator units', six 'output-tables' and one 'input table', six integrations may be carried out simultaneously and π_{n-2}, π_{n-4}, π_{n-6}, π_{n-8}, π_{n-10} and π_{n-12} produced from $\pi_n(\rho)$ in one 'run'.

REFERENCES

Campbell, G. A. (1923). *Bell. Syst. Tech. J.* January, 1923.
Hartley, H. O. (1938). *J.R. Statist. Soc. Suppl.* **5**, 80.
Hartree, D. R. (1935) *Nature, Lond.*, **135**, 940.
Newman, D. (1939). *Biometrika*, **31**, 20.
Pearson, E. S. & Hartley, H. O. (1943). *Biometrika*, **33**, 89.
Simon, L. E. (1941). *An Engineer's Manual of Statistical Methods.* New York: John Wiley and Sons, Inc.
Soper, H. E. (1921). *Tracts for Computers*, No. VII. Camb. Univ. Press.
Thompson, C. M. (1941). *Biometrika*, **32**, 151.
Wishart, J. (1927). *Biometrika*, **19**, 1.

† Note that with the recurrence process (2) only one numerical quadrature is required to produce $p_{n-2}(R)$ for all values of R, whilst the exact formula (1) necessitates a separate quadrature over the range 0 to ∞ for every single value of R.

MISCELLANEA

Note on the use of the tables of percentage points of the incomplete beta function to calculate small sample confidence intervals for a binomial p

BY HENRY SCHEFFÉ, *Princeton University*

One of many once tedious statistical computations which may now be made simply and directly by the use of Miss Thompson's recent tables (1941) is the calculation of confidence intervals for a binomial p when the sample size is small, so that one hesitates to use the normal approximation to the binomial distribution. The problem was treated by C. J. Clopper & E. S. Pearson (1934). In their article they included charts yielding confidence intervals for $\epsilon = 0{\cdot}01$ and $0{\cdot}05$, where $1-\epsilon$ denotes the confidence coefficient. With the new tables these cases as well as $\epsilon = 0{\cdot}02$, $0{\cdot}1$, $0{\cdot}2$, $0{\cdot}5$ are easily handled.

Let x be the number of 'successes' observed in a sample of n trials on a binomial population for which $E(x/n) = p$. Denote by $p_1(x) \leqq p \leqq p_2(x)$ a confidence interval for p with confidence coefficient $1-\epsilon$. From the work of Clopper & Pearson we find that $p_2(x)$ is determined by the equation

$$\sum_{j=0}^{x} {}_nC_j p_2^j (1-p_2)^{n-j} = \tfrac{1}{2}\epsilon \quad (x<n), \tag{1}$$

while $p_2(n) = 1$, and that $p_1(x)$ may then be found from

$$p_1(x) = 1 - p_2(n-x). \tag{2}$$

Karl Pearson (1924) showed how the left member of (1) may be evaluated in terms of the incomplete Beta function,

$$\sum_{j=0}^{x} {}_nC_j p^j (1-p)^{n-j} = I_{1-p}(n-x, x+1) \quad (x<n). \tag{3}$$

From (1), (2), (3) we deduce the following rule for calculating the confidence limits $p_1(x)$ and $p_2(x)$ from Miss Thompson's tables: Enter the table for the $100(\frac{1}{2}\epsilon)$ percentage point with $\nu_1 = 2(n-x+1)$, $\nu_2 = 2x$ to read $p_1(x)$ directly; in the same table subtract from unity the entry for $\nu_1 = 2(x+1)$, $\nu_2 = 2(n-x)$ to get $p_2(x)$. The exceptions to this rule occur for $p_1(0)$ and $p_2(n)$, to which we assign the values 0 and 1, respectively. If the percentage point for the desired ν_1, ν_2 is not tabulated, Hartley's *Methods of Interpolation* preceding the tables will yield it fairly quickly, especially since only two decimals will be wanted ordinarily.

REFERENCES

CLOPPER, C. J. & PEARSON, E. S. (1934). *Biometrika*, **26**, 404–13.
PEARSON, KARL (1924). *Biometrika*, **16**, 202–3.
THOMPSON, CATHERINE M. (1941). *Biometrika*, **32**, 151–81.

BOOKS RECEIVED FOR REVIEW

The Genetics of the Mouse. By HANS GRÜNEBERG. Cambridge: The University Press. 1943. Price 30*s*.

The Advanced Theory of Statistics. Vol. I. By MAURICE G. KENDALL. London: Charles Griffin and Co. Ltd. 1943. Price 42*s*.

The Statistical Study of Literary Vocabulary. By G. UDNY YULE. Cambridge: The University Press. 1944. Price 25*s*.

The Year Book of Labour Statistics. Seventh year of issue. 1942. Montreal: International Labour Office. 1943. Price 8*s*.

(1) *Table of Circular and Hyperbolic Tangents and Cotangents for Radian Arguments.* (2) *Table of the Bessel Functions* $J_0(z)$ *and* $J_1(z)$ *for Complex Arguments.* (3) *Tables of Lagrangian Interpolation Coefficients.* (4) *Table of Reciprocals of Integers from* 100,000 *through* 200,009. Tables prepared by the Mathematical Tables Project, Works Projects Administration of the Federal Works Agency, conducted under the sponsorship of the National Bureau of Standards and edited by LYMAN J. BRIGGS and ARNOLD N. LOWAN. New York: Columbia University Press. 1943–44. Price: (1)–(3) $5.00 each volume, (4) $4.00.

(*All Rights reserved*)

BIOMETRIKA. Vol. XXXIII, Part II

CONTENTS

PAGE

On autoregressive time series. By M. G. KENDALL . . . . . 105—122

Comparison of the concepts of efficiency and closeness for consistent estimates of a parameter. By R. C. GEARY . . . . . 123—128

The relation between measures of correlation in the universe of sample permutations. By H. E. DANIELS . . . . . 129—135

The growth, survival, wandering and variation of the long-tailed field mouse, *Apodemus sylvaticus*. By H. P. HACKER and H. S. PEARSON . . . . . 136—162

The control of industrial processes subject to trends in quality. By L. H. C. TIPPETT . . 163—172

Studentization, or the elimination of the standard deviation of the parent population from the random sample-distribution of statistics. By H. O. HARTLEY . . . . . 173—180

MISCELLANEA:

Note on the use of the tables of percentage points of the incomplete beta function to calculate small sample confidence intervals for a binomial p. By HENRY SCHEFFÉ 181

BOOKS RECEIVED FOR REVIEW 182

PRINTED IN ENGLAND BY W. LEWIS, M.A., AT THE CAMBRIDGE UNIVERSITY PRESS
Reprinted by Photo-Lithography by Warren & Son Ltd., Winchester.